Chapter 3 Convex Functions and Generalizations

Convex and concave functions have many special and important properties. For example, any local minimum of a convex function over a convex set is also a global minimum. In this chapter we introduce the important topics of convex and concave functions and develop some of their properties. As we shall learn in this and later chapters, these properties can be utilized in developing suitable optimality conditions and computational schemes for optimization problems that involve convex and concave functions.

Following is an outline of the chapter.

Section 3.1: Definitions and Basic Properties We introduce convex and concave functions and develop some of their basic properties. Continuity of convex functions is proved, and the concept of a directional derivative is introduced.

Section 3.2: Subgradients of Convex Functions A convex function has a convex epigraph and hence has a supporting hyperplane. This leads to the important notion of a subgradient of a convex function.

Section 3.3: Differentiable Convex Functions In this section we give some characterizations of differentiable convex functions. These are helpful tools for checking convexity of simple differentiable functions.

Section 3.4: Minima and Maxima of Convex Functions This section is important, since it deals with the questions of minimizing and maximizing a convex function over a convex set. A necessary and sufficient condition for a minimum is developed, and we provide a characterization for the set of alternative optimal solutions. We also show that the maximum occurs at an extreme point. This fact is particularly important if the convex set is polyhedral.

Section 3.5: Generalizations of Convex Functions Various relaxations of convexity and concavity are possible. We present quasiconvex and pseudoconvex functions and develop some of their properties. We then discuss various types of convexity at a point. These types of convexity are sometimes sufficient for optimality, as shown in Chapter 4. (This section can be omitted by beginning readers, and later references to generalized convexity properties can largely be substituted simply by convexity.)

3.1 Definitions and Basic Properties

In this section we deal with some basic properties of convex and concave functions. In particular, we investigate their continuity and differentiability properties.

3.1.1 Definition

Let $f: S \to R$, where S is a nonempty convex set in R^n. The function f is said to be *convex* on S if

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \le \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

for each $\mathbf{x}_1, \mathbf{x}_2 \in S$ and for each $\lambda \in (0, 1)$. The function f is called ***strictly convex*** on S if the above inequality is true as a strict inequality for each distinct $\mathbf{x}_1$ and $\mathbf{x}_2$ in S and for each $\lambda \in (0, 1)$. The function $f: S \to R$ is called *concave* (*strictly concave*) on S if $-f$ is convex (strictly convex) on S.

Now let us consider the geometric interpretation of convex and concave functions. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two distinct points in the domain of f, and consider the point $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, with $\lambda \in (0, 1)$. Note that $\lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$ gives the weighted average of $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$, while $f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2]$ gives the value of f at the point $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. So for a convex function f, the value of f at points on the line segment $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ is less than or equal to the height of the chord joining the points $[\mathbf{x}_1, f(\mathbf{x}_1)]$ and $[\mathbf{x}_2, f(\mathbf{x}_2)]$. For a concave function, the chord is (on or) below the function itself. Hence, a function is both convex and concave if and only if it is *affine*. Figure 3.1 shows some examples of convex and concave functions.

The following are some examples of convex functions. By taking the negatives of these functions, we get some examples of concave functions.

1. $f(x) = 3x + 4$.
2. $f(x) = |x|$.
3. $f(x) = x^2 - 2x$.
4. $f(x) = -x^{1/2}$ if $x \ge 0$.
5. $f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1x_2$.
6. $f(x_1, x_2, x_3) = x_1^4 + 2x_2^2 + 3x_3^2 - 4x_1 - 4x_2x_3$.

Note that in each of the above examples, except for Example 4, the function f is convex over R^n. In Example 4 the function is not defined for $x < 0$. One can readily construct examples of functions that are convex over a region but not over R^n. For instance, $f(x) = x^3$ is not convex over R but is convex over $S = \{x : x \ge 0\}$.

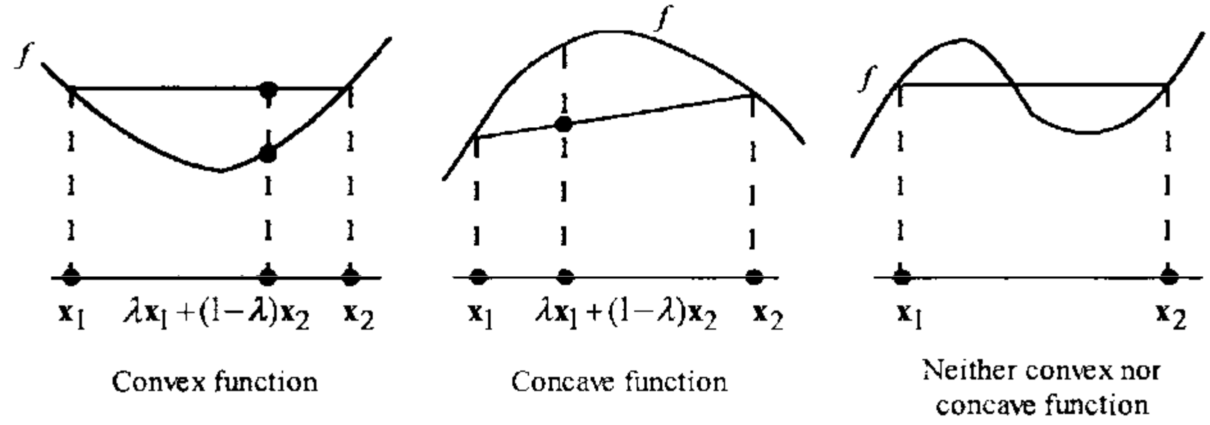


Figure 3.1 Convex and concave functions.

The examples above cite some arbitrary illustrative instances of convex functions. In contrast, we give below some particularly important instances of convex functions that arise very often in practice and that are useful to remember.

1 Let $f_1, f_2,\ldots, f_k$: $R^n \to R$ be convex functions. Then:

(a) $f(\mathbf{x}) = \sum_{j=1}^{k} \alpha_j f_j(\mathbf{x})$, where $\alpha_j > 0$ for $j = 1, 2,\ldots, k$ is a convex function (see Exercise 3.8).

(b) $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}),\ldots, f_k(\mathbf{x})\}$ is a convex function (see Exercise 3.9).

2. Suppose that g: $R^n \to R$ is a concave function. Let $S = \{\mathbf{x} : g(\mathbf{x}) > 0\}$, and define f: $S \to R$ as $f(\mathbf{x}) = 1/g(\mathbf{x})$. Then f is convex over S (see Exercise 3.11).

3. Let g: $R \to R$ be a nondecreasing, univariate, convex function, and let h: $R^n \to R$ be a convex function. Then the composite function f: $R^n \to R$ defined as $f(\mathbf{x}) = g[h(\mathbf{x})]$ is a convex function (see Exercise 3.10).

4. Let g: $R^m \to R$ be a convex function, and let $\mathbf{h}$: $R^n \to R^m$ be an affine function of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an $m \times 1$ vector. Then the composite function f: $R^n \to R$ defined as $f(\mathbf{x}) = g[\mathbf{h}(\mathbf{x})]$ is a convex function (see Exercise 3.16).

From now on, we concentrate on convex functions. Results for concave functions can be obtained easily by noting that f is concave if and only if $-f$ is convex.

Associated with a convex function f is the set $S_\alpha = \{\mathbf{x} \in S : f(\mathbf{x}) \le \alpha\}$, $\alpha \in R$, usually referred to as a *level set*. Sometimes this set is called a *lower-level set*, to differentiate it from the *upper-level set* $\{\mathbf{x} \in S : f(\mathbf{x}) \ge \alpha\}$, which has properties similar to these for concave functions. Lemma 3.1.2 shows that S_α is convex for each real number α. Hence, if g_i: $R^n \to R$ is convex for $i = 1,\ldots, m$, the set $\{\mathbf{x} : g_i(\mathbf{x}) \le 0, i = 1,\ldots, m\}$ is a convex set.

3.1.2 Lemma

Let S be a nonempty convex set in R^n, and let f: $S \to R$ be a convex function. Then the level set $S_\alpha = \{\mathbf{x} \in S : f(\mathbf{x}) \le \alpha\}$, where α is a real number, is a convex set.

Proof

Let $\mathbf{x}_1$, $\mathbf{x}_2 \in S_\alpha$. Thus, $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ and $f(\mathbf{x}_1) \le \alpha$ and $f(\mathbf{x}_2) \le \alpha$. Now let $\lambda \in (0, 1)$ and $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. By the convexity of S, we have that $\mathbf{x} \in S$. Furthermore, by the convexity of f,

$$f(\mathbf{x}) \le \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \le \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

Hence, $\mathbf{x} \in S_\alpha$, and therefore, S_α is convex.

Continuity of Convex Functions

An important property of convex and concave functions is that they are continuous on the interior of their domain. This fact is proved below.

3.1.3 Theorem

Let S be a nonempty convex set in R^n, and let f: $S \to R$ be convex. Then f is continuous on the interior of S.

Proof

Let $\bar{\mathbf{x}} \in \operatorname{int} S$. To prove continuity of f at $\bar{\mathbf{x}}$, we need to show that given $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|\mathbf{x} - \bar{\mathbf{x}}\| \le \delta$ implies that $|f(\mathbf{x}) - f(\bar{\mathbf{x}})| \le \varepsilon$. Since $\bar{\mathbf{x}} \in \operatorname{int} S$, there exists a $\delta' > 0$ such that $\|\mathbf{x} - \bar{\mathbf{x}}\| \le \delta'$ implies that $\mathbf{x} \in S$. Construct θ as follows.

$$\theta = \max_{1 \le i \le n}\{\max[f(\bar{\mathbf{x}} + \delta'\mathbf{e}_i) - f(\bar{\mathbf{x}}), f(\bar{\mathbf{x}} - \delta'\mathbf{e}_i) - f(\bar{\mathbf{x}})]\}, \tag{3.1}$$

where $\mathbf{e}_i$ is a vector of zeros except for a 1 at the ith position. Note that $0 \le \theta < \infty$. Let

$$\delta = \min\left(\frac{\delta'}{n}, \frac{\varepsilon\delta'}{n\theta}\right). \tag{3.2}$$

Choose an x with $\|\mathbf{x}-\bar{\mathbf{x}}\| \le \delta$. If $x_i - \bar{x}_i \ge 0$, let $\mathbf{z}_i = \delta' \mathbf{e}_i$; otherwise, let $\mathbf{z}_i = -\delta' \mathbf{e}_i$. Then $\mathbf{x}-\bar{\mathbf{x}} = \sum_{i=1}^{n} \alpha_i \mathbf{z}_i$, where $\alpha_i \ge 0$ for $i = 1,\ldots, n$. Furthermore,

$$\|\mathbf{x}-\bar{\mathbf{x}}\| = \delta' \left(\sum_{i=1}^{n} \alpha_i^2 \right)^{1/2}. \tag{3.3}$$

From (3.2), and since $\|\mathbf{x}-\bar{\mathbf{x}}\| \le \delta$, it follows that $\alpha_i \le 1/n$ for $i = 1,\ldots, n$. Hence, by the convexity of f, and since $0 \le n\alpha_i \le 1$, we get

$$\begin{aligned}
f(\mathbf{x}) = f\left(\bar{\mathbf{x}} + \sum_{i=1}^{n} \alpha_i \mathbf{z}_i \right) &= f\left[\frac{1}{n} \sum_{i=1}^{n} (\bar{\mathbf{x}} + n\alpha_i \mathbf{z}_i) \right] \\
&\le \frac{1}{n} \sum_{i=1}^{n} f(\bar{\mathbf{x}} + n\alpha_i \mathbf{z}_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} f[(1-n\alpha_i)\bar{\mathbf{x}} + n\alpha_i(\bar{\mathbf{x}} + \mathbf{z}_i)] \\
&\le \frac{1}{n} \sum_{i=1}^{n} [(1-n\alpha_i) f(\bar{\mathbf{x}}) + n\alpha_i f(\bar{\mathbf{x}} + \mathbf{z}_i)].
\end{aligned}$$

Therefore, $f(\mathbf{x}) - f(\bar{\mathbf{x}}) \le \sum_{i=1}^{n} \alpha_i [f(\bar{\mathbf{x}} + \mathbf{z}_i) - f(\bar{\mathbf{x}})]$. From (3.1) it is obvious that $f(\bar{\mathbf{x}} + \mathbf{z}_i) - f(\bar{\mathbf{x}}) \le \theta$ for each i; and since $\alpha_i \ge 0$, it follows that

$$f(\mathbf{x}) - f(\bar{\mathbf{x}}) \le \theta \sum_{i=1}^{n} \alpha_i. \tag{3.4}$$

Noting (3.3) and (3.2), it follows that $\alpha_i \le \varepsilon / n\theta$, and (3.4) implies that $f(\mathbf{x}) - f(\bar{\mathbf{x}}) \le \varepsilon$. So far, we have shown that $\|\mathbf{x}-\bar{\mathbf{x}}\| \le \delta$ implies that $f(\mathbf{x}) - f(\bar{\mathbf{x}}) \le \varepsilon$. By definition, this establishes the *upper semicontinuity* of f at $\bar{\mathbf{x}}$. To complete the proof, we need to establish the *lower semicontinuity* of f at $\bar{\mathbf{x}}$ as well, that is, to show that $f(\bar{\mathbf{x}}) - f(\mathbf{x}) \le \varepsilon$. Let $\mathbf{y} = 2\bar{\mathbf{x}} - \mathbf{x}$ and note that $\|\mathbf{y}-\bar{\mathbf{x}}\| \le \delta$. Therefore, as above,

$$f(\mathbf{y}) - f(\bar{\mathbf{x}}) \le \varepsilon. \tag{3.5}$$

But $\bar{\mathbf{x}} = (1/2)\mathbf{y} + (1/2)\mathbf{x}$, and by the convexity of f, we have

$$f(\bar{\mathbf{x}}) \le (1/2) f(\mathbf{y}) + (1/2) f(\mathbf{x}). \tag{3.6}$$

Combining (3.5) and (3.6) above, it follows that $f(\bar{\mathbf{x}}) - f(\mathbf{x}) \le \varepsilon$, and the proof is complete.

Note that convex and concave functions may not be continuous everywhere. However, by Theorem 3.1.3, points of discontinuity only are allowed at the boundary of S, as illustrated by the following convex function defined on $S = \{x : -1 \le x \le 1\}$:

$$f(x) = \begin{cases} x^2 & \text{for } |x| < 1 \\ 2 & \text{for } |x| = 1. \end{cases}$$

Directional Derivative of Convex Functions

The concept of directional derivatives is particularly useful in the motivation and development of some optimality criteria and computational procedures in nonlinear programming, where one is interested in finding a direction along which the function decreases or increases.

3.1.4 Definition

Let S be a nonempty set in R^n, and let $f: S \to R$. Let $\bar{\mathbf{x}} \in S$ and $\mathbf{d}$ be a nonzero vector such that $\bar{\mathbf{x}} + \lambda\mathbf{d} \in S$ for $\lambda > 0$ and sufficiently small. The *directional derivative* of f at $\bar{\mathbf{x}}$ along the vector $\mathbf{d}$, denoted by $f'(\bar{\mathbf{x}};\mathbf{d})$, is given by the following limit if it exists:

$$f'(\bar{\mathbf{x}};\mathbf{d}) = \lim_{\lambda \to 0^+} \frac{f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}})}{\lambda}.$$

In particular, the limit in Definition 3.1.4 exists for globally defined convex and concave functions as shown below. As evident from the proof of the following lemma, if $f: S \to R$ is convex on S, the limit exists if $\bar{\mathbf{x}} \in \text{int } S$, but might be $-\infty$ if $\bar{\mathbf{x}} \in \partial S$, even if f is continuous at $\bar{\mathbf{x}}$, as seen in Figure 3.2.

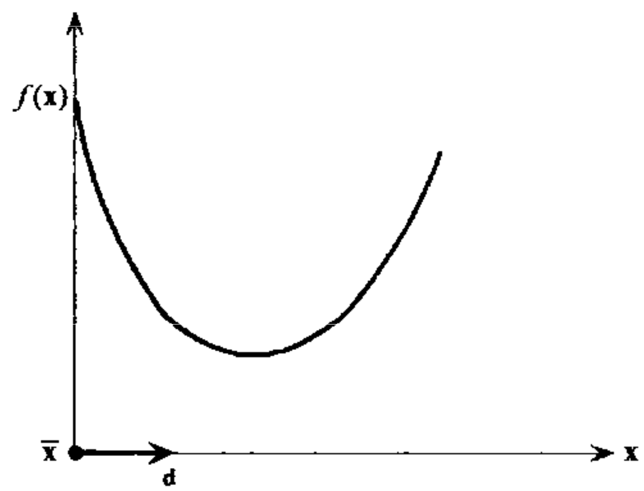


Figure 3.2 Nonexistence of the directional derivative of f at $\bar{\mathbf{x}}$ in the direction d.

3.1.5 Lemma

Let $f: R^n \to R$ be a convex function. Consider any point $\bar{x} \in R^n$ and a nonzero direction $d \in R^n$. Then the directional derivative $f'(\bar{x};d)$, of f at $\bar{x}$ in the direction **d**, exists.

Proof

Let $\lambda_2 > \lambda_1 > 0$. Noting the convexity of f, we have

$$f(\bar{x}+\lambda_1 d) = f\left[\frac{\lambda_1}{\lambda_2}(\bar{x}+\lambda_2 d)+\left(1-\frac{\lambda_1}{\lambda_2}\right)\bar{x}\right]$$
$$\leq \frac{\lambda_1}{\lambda_2} f(\bar{x}+\lambda_2 d)+\left(1-\frac{\lambda_1}{\lambda_2}\right) f(\bar{x}).$$

This inequality implies that

$$\frac{f(\bar{x}+\lambda_1 d)-f(\bar{x})}{\lambda_1} \leq \frac{f(\bar{x}+\lambda_2 d)-f(\bar{x})}{\lambda_2}.$$

Thus, the difference quotient $[f(\bar{x}+\lambda d)-f(\bar{x})]/\lambda$ is monotone decreasing (nonincreasing) as $\lambda \to 0^+$.

Now, given any $\lambda \geq 0$, we also have, by the convexity of f, that

$$f(\bar{x}) = f\left[\frac{\lambda}{1+\lambda}(\bar{x}-d)+\frac{1}{1+\lambda}(\bar{x}+\lambda d)\right]$$
$$\leq \frac{\lambda}{1+\lambda} f(\bar{x}-d)+\frac{1}{1+\lambda} f(\bar{x}+\lambda d).$$

So

$$\frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda} \geq f(\bar{x})-f(\bar{x}-d).$$

Hence, the monotone decreasing sequence of values $[f(\bar{x}+\lambda d)-f(\bar{x})]/\lambda$, as $\lambda \to 0^+$, is bounded from below by the constant $f(\bar{x})-f(\bar{x}-d)$. Hence, the limit in the theorem exists and is given by

$$\lim_{\lambda \to 0^+} \frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda} = \inf_{\lambda > 0} \frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda}.$$

3.2 Subgradients of Convex Functions

In this section, we introduce the important concept of subgradients of convex and concave functions via supporting hyperplanes to the epigraphs of convex functions and to the hypographs of concave functions.

Epigraph and Hypograph of a Function

A function f on S can be fully described by the set $\{[\mathbf{x}, f(\mathbf{x})] : \mathbf{x} \in S\} \subset R^{n+1}$, which is referred to as the *graph* of the function. One can construct two sets that are related to the graph of f: the epigraph, which consists of points above the graph of f, and the hypograph, which consists of points below the graph of f. These notions are clarified in Definition 3.2.1.

3.2.1 Definition

Let S be a nonempty set in R^n, and let $f: S \to R$. The *epigraph* of f, denoted by epi f, is a subset of R^{n+1} defined by

$$\{(\mathbf{x}, y) : \mathbf{x} \in S,\ \ y \in R,\ \ y \geq f(\mathbf{x})\}.$$

The *hypograph* of f, denoted by hyp f, is a subset of R^{n+1} defined by

$$\{(\mathbf{x}, y) : \mathbf{x} \in S,\ \ y \in R,\ \ y \leq f(\mathbf{x})\}.$$

Figure 3.3 illustrates the epigraphs and hypographs of several functions. In Figure 3.3*a*, neither the epigraph nor the hypograph of f is a convex set. But in Figure 3.3*b* and *c*, respectively, the epigraph and hypograph of f are convex sets. It turns out that a function is convex if and only if its epigraph is a convex set and, equivalently, that a function is concave if and only if its hypograph is a convex set.

3.2.2 Theorem

Let S be a nonempty convex set in R^n, and let $f: S \to R$. Then f is convex if and only if epi f is a convex set.

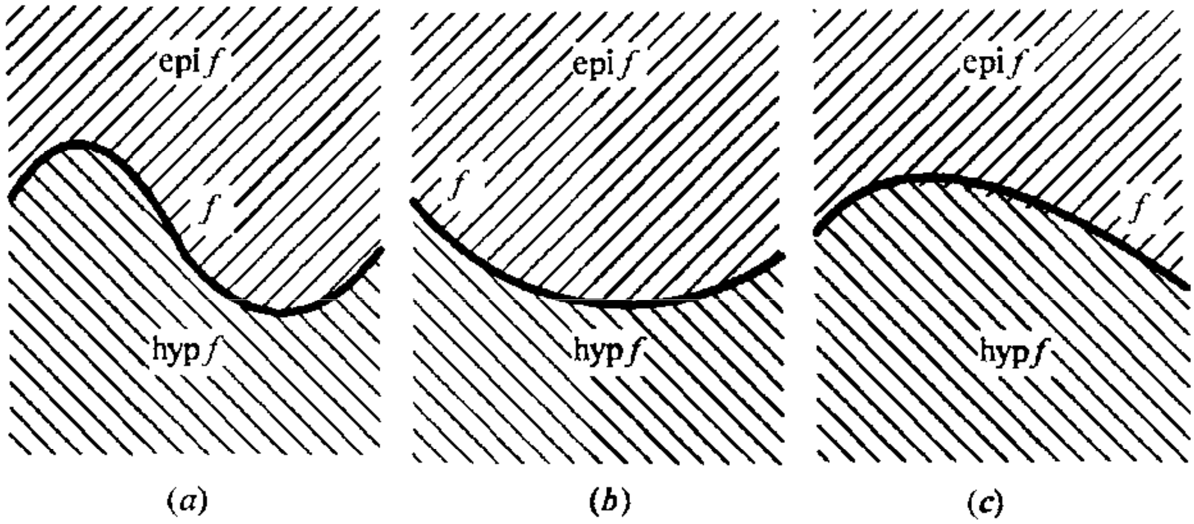


Figure 3.3 Epigraphs and hypographs.

Proof

Assume that f is convex, and let $(\mathbf{x}_1, y_1)$ and $(\mathbf{x}_1, y_2) \in \text{epi} f$; that is, $\mathbf{x}_1$, $\mathbf{x}_2 \in S$, $y_1 \ge f(\mathbf{x}_1)$, and $y_2 \ge f(\mathbf{x}_2)$. Let $\lambda \in (0, 1)$. Then

$$\lambda y_1 + (1-\lambda) y_2 \ge \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \ge f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2),$$

where the last inequality follows by the convexity of f. Note that $\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in S$. Thus, $[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda y_1 + (1-\lambda) y_2] \in \text{epi} f$, and hence epi f is convex. Conversely, assume that epi f is convex, and let $\mathbf{x}_1$, $\mathbf{x}_2 \in S$. Then $[\mathbf{x}_1, f(\mathbf{x}_1)]$ and $[\mathbf{x}_2, f(\mathbf{x}_2)]$ belong to epi f, and by the convexity of epi f, we must have

$$[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)] \in \text{epi} f \qquad \text{for } \lambda \in (0, 1).$$

In other words, $\lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \ge f[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2]$ for each $\lambda \in (0, 1)$; that is, f is convex. This completes the proof.

Theorem 3.2.2 can be used to verify the convexity or concavity of a given function f. Making use of this result, it is clear that the functions illustrated in Figure 3.3 are (*a*) neither convex nor concave, (*b*) convex, and (*c*) concave.

Since the epigraph of a convex function and the hypograph of a concave function are convex sets, they have supporting hyperplanes at points of their boundary. These supporting hyperplanes lead to the notion of subgradients, which is defined below.

3.2.3 Definition

Let S be a nonempty convex set in R^n, and let $f: S \to R$ be convex. Then $\boldsymbol{\xi}$ is called a *subgradient of f* at $\bar{\mathbf{x}} \in S$ if

$$f(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \boldsymbol{\xi}^t (\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for all } \mathbf{x} \in S.$$

Similarly, let $f: S \to R$ be concave. Then $\boldsymbol{\xi}$ is called a *subgradient* of f at $\bar{\mathbf{x}} \in S$ if

$$f(\mathbf{x}) \le f(\bar{\mathbf{x}}) + \boldsymbol{\xi}^t (\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for all } \mathbf{x} \in S.$$

From Definition 3.2.3 it follows immediately that the collection of subgradients of f at $\bar{\mathbf{x}}$ (known as the *subdifferential* of f at $\bar{\mathbf{x}}$) is a convex set. Figure 3.4 shows examples of subgradients of convex and concave functions. From the figure we see that the function $f(\bar{\mathbf{x}}) + \boldsymbol{\xi}^t (\mathbf{x} - \bar{\mathbf{x}})$ corresponds to a supporting hyperplane of the epigraph or the hypograph of the function f. The subgradient vector $\boldsymbol{\xi}$ corresponds to the slope of the supporting hyperplane.

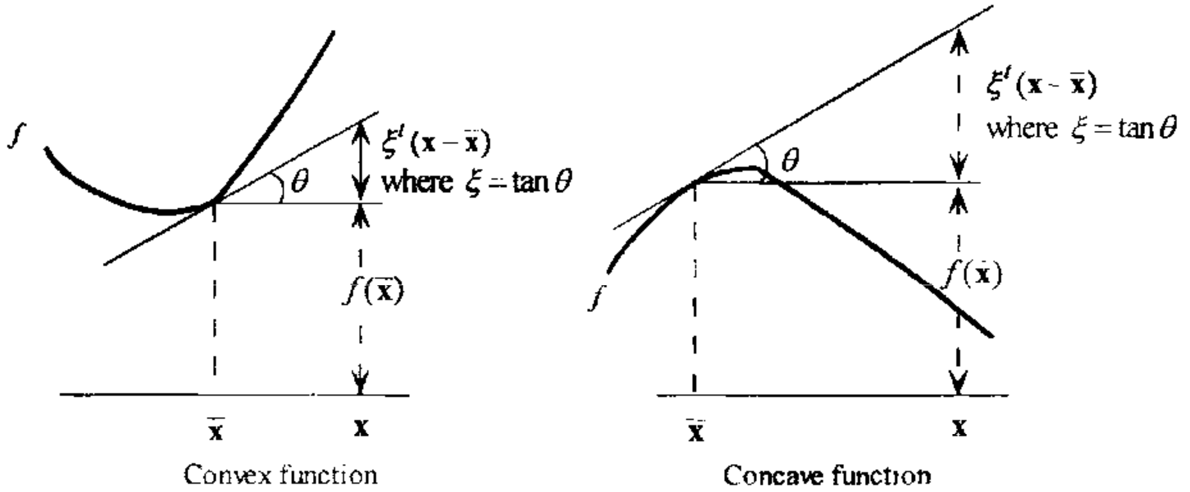


Figure 3.4 Geometric interpretation of subgradients.

3.2.4 Example

Let $f(x) = \min\{f_1(x), f_2(x)\}$, where f_1 and f_2 are as defined below:

$$f_1(x) = 4 - |x|, \qquad x \in R$$

$$f_2(x) = 4 - (x-2)^2, \qquad x \in R.$$

Since $f_2(x) \ge f_1(x)$ for $1 \le x \le 4$, f can be represented as follows:

$$f(x) = \begin{cases} 4 - x, & 1 \le x \le 4 \\ 4 - (x-2)^2, & \text{otherwise.} \end{cases}$$

In Figure 3.5 the concave function f is shown in dark lines. Note that $\xi = -1$ is the slope and hence the subgradient of f at any point x in the open interval $(1, 4)$. If $x < 1$ or $x > 4$, $\xi = -2(x-2)$ is the unique subgradient of f. At the points $x = 1$ and $x = 4$, the subgradients are not unique because many supporting hyperplanes exist. At $x = 1$, the family of subgradients is characterized by $\lambda\nabla f_1(1) + (1-\lambda)\nabla f_2(1) = \lambda(-1) + (1-\lambda)(2) = 2 - 3\lambda$ for $\lambda \in [0, 1]$. In other words, any ξ in the interval $[-1, 2]$ is a subgradient of f at $x = 1$, and this corresponds to the slopes of the family of supporting hyperplanes of f at $x = 1$. At $x = 4$, the family of subgradients is characterized by $\lambda\nabla f_1(4) + (1-\lambda)\nabla f_2(4) = \lambda(-1) + (1-\lambda)(-4) = -4 + 3\lambda$ for $\lambda \in [0, 1]$. In other words, any ξ in the interval $[-4, -1]$ is a subgradient of f at $x = 4$. Exercise 3.27 addresses the general characterization of subgradients of functions of the form $f(\mathbf{x}) = \min\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$.

The following theorem shows that every convex or concave function has at least one subgradient at points in the interior of its domain. The proof relies on the fact that a convex set has a supporting hyperplane at points of the boundary.

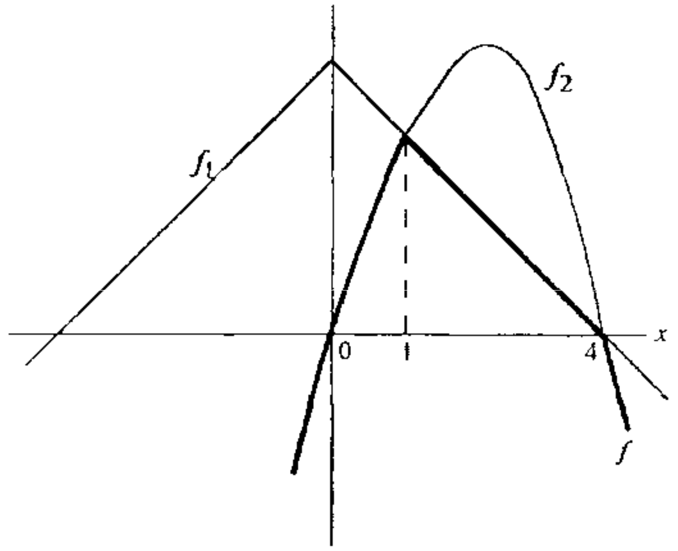


Figure 3.5 Setup for Example 3.2.4.

3.2.5 Theorem

Let S be a nonempty convex set in R^n, and let $f: S \to R$ be convex. Then for $\bar{\mathbf{x}} \in \operatorname{int} S$, there exists a vector $\boldsymbol{\xi}$ such that the hyperplane

$$H = \{(\mathbf{x}, y) : y = f(\bar{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \bar{\mathbf{x}})\}$$

supports epi f at $[\bar{\mathbf{x}}, f(\bar{\mathbf{x}})]$. In particular,

$$f(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for each } \mathbf{x} \in S;$$

that is, $\boldsymbol{\xi}$ is a subgradient of f at $\bar{\mathbf{x}}$.

Proof

By Theorem 3.2.2, epi f is convex. Noting that $[\bar{\mathbf{x}}, f(\bar{\mathbf{x}})]$ belongs to the boundary of epi f, by Theorem 2.4.7 there exists a nonzero vector $(\boldsymbol{\xi}_0, \mu) \in R^n \times R$ such that

$$\boldsymbol{\xi}_0^t(\mathbf{x} - \bar{\mathbf{x}}) + \mu[y - f(\bar{\mathbf{x}})] \le 0 \qquad \text{for all } (\mathbf{x}, y) \in \text{ epi} f. \tag{3.7}$$

Note that μ is not positive, because otherwise, inequality (3.7) will be contradicted by choosing y sufficiently large. We now show that $\mu < 0$. By contradiction, suppose that $\mu = 0$. Then $\boldsymbol{\xi}_0^t(\mathbf{x} - \bar{\mathbf{x}}) \le 0$ for all $\mathbf{x} \in S$. Since $\bar{\mathbf{x}} \in \operatorname{int} S$, there exists a $\lambda > 0$ such that $\bar{\mathbf{x}} + \lambda\boldsymbol{\xi}_0 \in S$ and hence $\lambda\boldsymbol{\xi}_0^t\boldsymbol{\xi}_0 \le 0$. This implies that $\boldsymbol{\xi}_0 = \mathbf{0}$ and $(\boldsymbol{\xi}_0, \mu) = (\mathbf{0}, 0)$, contradicting the fact that $(\boldsymbol{\xi}_0, \mu)$ is a nonzero vector. Therefore, $\mu < 0$. Denoting $\boldsymbol{\xi}_0/|\mu|$ by $\boldsymbol{\xi}$ and dividing the inequality in (3.7) by $|\mu|$, we get

$$\boldsymbol{\xi}^t(\mathbf{x} - \bar{\mathbf{x}}) - y + f(\bar{\mathbf{x}}) \le 0 \qquad \text{for all } (\mathbf{x}, y) \in \text{epi} f. \tag{3.8}$$

In particular, the hyperplane $H = \{(\mathbf{x}, y) : y = f(\overline{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \overline{\mathbf{x}})\}$ supports epi f at $[\overline{\mathbf{x}}, f(\overline{\mathbf{x}})]$. By letting $y = f(\overline{\mathbf{x}})$ in (3.8), we get $f(\mathbf{x}) \ge f(\overline{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \overline{\mathbf{x}})$ for all $\mathbf{x} \in S$, and the proof is complete.

Corollary

Let S be a nonempty convex set in R^n, and let $f: S \to R$ be strictly convex. Then for $\overline{\mathbf{x}} \in \text{int}\, S$ there exists a vector $\boldsymbol{\xi}$ such that

$$f(\mathbf{x}) > f(\overline{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \overline{\mathbf{x}}) \qquad \text{for all } \mathbf{x} \in S,\ \mathbf{x} \ne \overline{\mathbf{x}}.$$

Proof

By Theorem 3.2.5 there exists a vector $\boldsymbol{\xi}$ such that

$$f(\mathbf{x}) \ge f(\overline{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \overline{\mathbf{x}}) \qquad \text{for all } \mathbf{x} \in S. \tag{3.9}$$

By contradiction, suppose that there is an $\hat{\mathbf{x}} \ne \overline{\mathbf{x}}$ such that $f(\hat{\mathbf{x}}) = f(\overline{\mathbf{x}}) + \boldsymbol{\xi}^t(\hat{\mathbf{x}} - \overline{\mathbf{x}})$. Then, by the strict convexity of f for $\lambda \in (0, 1)$, we get

$$f[\lambda\overline{\mathbf{x}} + (1-\lambda)\hat{\mathbf{x}}] < \lambda f(\overline{\mathbf{x}}) + (1-\lambda) f(\hat{\mathbf{x}}) = f(\overline{\mathbf{x}}) + (1-\lambda)\boldsymbol{\xi}^t(\hat{\mathbf{x}} - \overline{\mathbf{x}}). \tag{3.10}$$

But letting $\mathbf{x} = \lambda\overline{\mathbf{x}} + (1-\lambda)\hat{\mathbf{x}}$ in (3.9), we must have

$$f[\lambda\overline{\mathbf{x}} + (1-\lambda)\hat{\mathbf{x}}] \ge f(\overline{\mathbf{x}}) + (1-\lambda)\boldsymbol{\xi}^t(\hat{\mathbf{x}} - \overline{\mathbf{x}}),$$

contradicting (3.10). This proves the corollary.

The converse of Theorem 3.2.5 is not true in general. In other words, if corresponding to each point $\overline{\mathbf{x}} \in \text{int}\, S$ there is a subgradient of f, then f is not necessarily a convex function. To illustrate, consider the following example, where f is defined on $S = \{(x_1, x_2) : 0 \le x_1, x_2 \le 1\}$:

$$f(x_1, x_2) = \begin{cases} 0, & 0 \le x_1 \le 1, \quad 0 < x_2 \le 1 \\ \dfrac{1}{4} - \left(x_1 - \dfrac{1}{2}\right)^2, & 0 \le x_1 \le 1, \quad x_2 = 0. \end{cases}$$

For each point in the interior of the domain, the zero vector is a subgradient of f. However, f is not convex on S since epi f is clearly not a convex set. However, as the following theorem shows, f is indeed convex on int S.

3.2.6 Theorem

Let S be a nonempty convex set in R^n, and let $f: S \to R$. Suppose that for each point $\overline{\mathbf{x}} \in \text{int}\, S$ there exists a subgradient vector $\boldsymbol{\xi}$ such that

$$f(\mathbf{x}) \geq f(\overline{\mathbf{x}}) + \xi^t(\mathbf{x} - \overline{\mathbf{x}}) \qquad \text{for each } \mathbf{x} \in S.$$

Then, f is convex on int S.

Proof

Let $\mathbf{x}_1$, $\mathbf{x}_2 \in \operatorname{int} S$, and let $\lambda \in (0, 1)$. By Corollary 1 to Theorem 2.2.2, int S is convex, and we must have $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \operatorname{int} S$. By assumption, there exists a subgradient ξ of f at $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. In particular, the following two inequalities hold true:

$$f(\mathbf{x}_1) \geq f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] + (1-\lambda)\xi^t(\mathbf{x}_1 - \mathbf{x}_2)$$

$$f(\mathbf{x}_2) \geq f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] + \lambda\xi^t(\mathbf{x}_2 - \mathbf{x}_1).$$

Multiplying the above two inequalities by λ and $(1 - \lambda)$, respectively, and adding, we obtain

$$\lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \geq f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2],$$

and the result follows.

3.3 Differentiable Convex Functions

We now focus on differentiable convex and concave functions. First, consider the following definition of differentiability.

3.3.1 Definition

Let S be a nonempty set in R^n, and let f: $S \to R$. Then f is said to be *differentiable* at $\overline{\mathbf{x}} \in$ int S if there exist a vector $\nabla f(\overline{\mathbf{x}})$, called the *gradient vector*, and a function α: $R^n \to R$ such that

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + \nabla f(\overline{\mathbf{x}})^t(\mathbf{x} - \overline{\mathbf{x}}) + \|\mathbf{x} - \overline{\mathbf{x}}\|\alpha(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) \qquad \text{for each } \mathbf{x} \in S,$$

where $\lim_{\mathbf{x} \to \overline{\mathbf{x}}} \alpha(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}) = 0$. The function f is said to be differentiable on the open set $S' \subseteq S$ if it is differentiable at each point in S'. The representation of f above is called a *first-order* (*Taylor series*) *expansion* of f at (or about) the point $\overline{\mathbf{x}}$; and without the implicitly defined *remainder term* involving the function α, the resulting representation is called a *first-order* (*Taylor series*) *approximation* of f at (or about) the point $\overline{\mathbf{x}}$.

Note that if f is differentiable at $\overline{\mathbf{x}}$, there could only be one gradient vector, and this vector is given by

$$\nabla f(\overline{\mathbf{x}}) = \left(\frac{\partial f(\overline{\mathbf{x}})}{\partial x_1}, \ldots, \frac{\partial f(\overline{\mathbf{x}})}{\partial x_n} \right)^t \equiv (f_1(\overline{\mathbf{x}}), \ldots, f_n(\overline{\mathbf{x}}))^t,$$

where $f_i(\overline{\mathbf{x}}) \equiv \partial f(\overline{\mathbf{x}})/\partial x_i$ is the partial derivative of f with respect to x_i at $\overline{\mathbf{x}}$ (see Exercise 3.36, and review Appendix A.4).

The following lemma shows that a differentiable convex function has only one subgradient, the gradient vector. Hence, the results of the preceding section can easily be specialized to the differentiable case, in which the gradient vector replaces subgradients.

3.3.2 Lemma

Let S be a nonempty convex set in R^n, and let $f: S \to R$ be convex. Suppose that f is differentiable at $\overline{\mathbf{x}} \in \operatorname{int} S$. Then the collection of subgradients of f at $\overline{\mathbf{x}}$ is the singleton set $\{\nabla f(\overline{\mathbf{x}})\}$.

Proof

By Theorem 3.2.5, the set of subgradients of f at $\overline{\mathbf{x}}$ is not empty. Now, let $\boldsymbol{\xi}$ be a subgradient of f at $\overline{\mathbf{x}}$. As a result of Theorem 3.2.5 and the differentiability of f at $\overline{\mathbf{x}}$, for any vector $\mathbf{d}$ and for λ sufficiently small, we get

$$f(\overline{\mathbf{x}}+\lambda\mathbf{d}) \ge f(\overline{\mathbf{x}}) + \lambda\boldsymbol{\xi}^t\mathbf{d}$$

$$f(\overline{\mathbf{x}}+\lambda\mathbf{d}) = f(\overline{\mathbf{x}}) + \lambda\nabla f(\overline{\mathbf{x}})^t\mathbf{d} + \lambda\|\mathbf{d}\|\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}).$$

Subtracting the equation from the inequality, we obtain

$$0 \ge \lambda[\boldsymbol{\xi} - \nabla f(\overline{\mathbf{x}})]^t\mathbf{d} - \lambda\|\mathbf{d}\|\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}).$$

If we divide by $\lambda > 0$ and let $\lambda \to 0^+$, it follows that $[\boldsymbol{\xi} - \nabla f(\overline{\mathbf{x}})]^t\mathbf{d} \le 0$. Choosing $\mathbf{d} = \boldsymbol{\xi} - \nabla f(\overline{\mathbf{x}})$, the last inequality implies that $\boldsymbol{\xi} = \nabla f(\overline{\mathbf{x}})$. This completes the proof.

In the light of Lemma 3.3.2, we give the following important characterization of differentiable convex functions. The proof is immediate from Theorems 3.2.5 and 3.2.6 and Lemma 3.3.2.

3.3.3 Theorem

Let S be a nonempty open convex set in R^n, and let $f: S \to R$ be differentiable on S. Then f is convex if and only if for any $\overline{\mathbf{x}} \in S$, we have

$$f(\mathbf{x}) \ge f(\overline{\mathbf{x}}) + \nabla f(\overline{\mathbf{x}})^t(\mathbf{x} - \overline{\mathbf{x}}) \qquad \text{for each } \mathbf{x} \in S.$$

Similarly, f is strictly convex if and only if for each $\overline{\mathbf{x}} \in S$, we have

$$f(\mathbf{x}) > f(\overline{\mathbf{x}}) + \nabla f(\overline{\mathbf{x}})^t(\mathbf{x} - \overline{\mathbf{x}}) \qquad \text{for each } \mathbf{x} \ne \overline{\mathbf{x}} \text{ in } S.$$

There are two evident implications of the above result that find use in various contexts. The first is that if we have an optimization problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$, where f is a convex function, then given any point $\overline{\mathbf{x}}$, the affine function $f(\bar{\mathbf{x}})+\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})$ bounds f from below. Hence, the minimum of $f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})$ over X (or over a relaxation of X) yields a lower bound on the optimum value of the given optimization problem, which can prove to be useful in an algorithmic approach. A second point in the same spirit is that this affine bounding function can be used to derive polyhedral outer approximations. For example, consider the set $X = \{\mathbf{x} : g_i(\mathbf{x}) \le 0,\ i = 1,\dots, m\}$, where g_i is a convex function for each $i = 1,\dots, m$. Given any point $\bar{\mathbf{x}}$, construct the polyhedral set $\bar{X} = \{\mathbf{x} : g_i(\bar{\mathbf{x}})+\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \le 0,\ i = 1,\dots, m\}$. Note that the polyhedral set $\bar{X}$ contains X and, hence, affords an *outer linearization* of this set, since for any $\mathbf{x} \in X$, we have $0 \ge g_i(\mathbf{x}) \ge g_i(\bar{\mathbf{x}})+\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})$ for $i = 1,\dots, m$ by Theorem 3.3.3. Such representations play a central role in many successive approximation algorithms for various nonlinear optimization problems.

The following theorem gives another necessary and sufficient characterization of differentiable convex functions. For a function of one variable, the characterization reduces to the slope being nondecreasing.

3.3.4 Theorem

Let S be a nonempty open convex set in R^n and let $f: S \to R$ be differentiable on S. Then f is convex if and only if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$ we have

$$[\nabla f(\mathbf{x}_2)-\nabla f(\mathbf{x}_1)]^t(\mathbf{x}_2-\mathbf{x}_1) \ge 0.$$

Similarly, f is strictly convex if and only if, for each distinct $\mathbf{x}_1, \mathbf{x}_2 \in S$, we have

$$[\nabla f(\mathbf{x}_2)-\nabla f(\mathbf{x}_1)]^t(\mathbf{x}_2-\mathbf{x}_1) > 0.$$

Proof

Assume that f is convex, and let $\mathbf{x}_1, \mathbf{x}_2 \in S$. By Theorem 3.3.3 we have

$$f(\mathbf{x}_1) \ge f(\mathbf{x}_2)+\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1-\mathbf{x}_2)$$

$$f(\mathbf{x}_2) \ge f(\mathbf{x}_1)+\nabla f(\mathbf{x}_1)^t(\mathbf{x}_2-\mathbf{x}_1).$$

Adding the two inequalities, we get $[\nabla f(\mathbf{x}_2)-\nabla f(\mathbf{x}_1)]^t(\mathbf{x}_2-\mathbf{x}_1) \ge 0$. To show the converse, let $\mathbf{x}_1, \mathbf{x}_2 \in S$. By the mean value theorem,

$$f(\mathbf{x}_2)-f(\mathbf{x}_1) = \nabla f(\mathbf{x})^t(\mathbf{x}_2-\mathbf{x}_1), \tag{3.11}$$

where $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\lambda \in (0, 1)$. By assumption, $[\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}_1)]^t$ $(\mathbf{x}-\mathbf{x}_1) \ge 0$; that is, $(1-\lambda)[\nabla f(\mathbf{x})-\nabla f(\mathbf{x}_1)]^t(\mathbf{x}_2-\mathbf{x}_1) \ge 0$. This implies that

$\nabla f(\mathbf{x})^t(\mathbf{x}_2-\mathbf{x}_1)\geq\nabla f(\mathbf{x}_1)^t(\mathbf{x}_2-\mathbf{x}_1)$. By (3.11) we get $f(\mathbf{x}_2)\geq f(\mathbf{x}_1)+\nabla f(\mathbf{x}_1)^t(\mathbf{x}_2-\mathbf{x}_1)$, so by Theorem 3.3.3, f is convex. The strict case is similar and the proof is complete.

Even though Theorems 3.3.3 and 3.3.4 provide necessary and sufficient characterizations of convex functions, checking these conditions is difficult from a computational standpoint. A simple and more manageable characterization, at least for quadratic functions, can be obtained, provided that the function is twice differentiable.

Twice Differentiable Convex and Concave Functions

A function f that is differentiable at $\overline{\mathbf{x}}$ is said to be twice differentiable at $\overline{\mathbf{x}}$ if the *second-order (Taylor series) expansion* representation of Definition 3.3.5 exists.

3.3.5 Definition

Let S be a nonempty set in R^n, and let $f: S \to R$. Then f is said to be *twice differentiable* at $\overline{\mathbf{x}} \in \operatorname{int} S$ if there exist a vector $\nabla f(\overline{\mathbf{x}})$, and an $n \times n$ symmetric matrix $\mathbf{H}(\overline{\mathbf{x}})$, called the *Hessian matrix*, and a function $\alpha: R^n \to R$ such that

$$f(\mathbf{x})=f(\overline{\mathbf{x}})+\nabla f(\mathbf{x})^t(\mathbf{x}-\overline{\mathbf{x}})+\frac{1}{2}(\mathbf{x}-\overline{\mathbf{x}})^t\mathbf{H}(\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})+\|\mathbf{x}-\overline{\mathbf{x}}\|^2\alpha(\overline{\mathbf{x}};\mathbf{x}-\overline{\mathbf{x}})$$

for each $\mathbf{x} \in S$, where $\lim_{\mathbf{x}\to\overline{\mathbf{x}}}\alpha(\overline{\mathbf{x}};\mathbf{x}-\overline{\mathbf{x}})=0$. The function f is said to be twice differentiable on the open set $S' \subseteq S$ if it is twice differentiable at each point in S'.

It may be noted that for twice differentiable functions, the Hessian matrix $\mathbf{H}(\overline{\mathbf{x}})$ is comprised of the second-order partial derivatives $f_{ij}(\overline{\mathbf{x}})\equiv \partial^2 f(\overline{\mathbf{x}})/\partial x_i\,\partial x_j$ for $i=1,\ldots,n$, $j=1,\ldots,n$, and is given as follows:

$$\mathbf{H}(\overline{\mathbf{x}})=\begin{bmatrix} f_{11}(\overline{\mathbf{x}}) & f_{12}(\overline{\mathbf{x}}) & \cdots & f_{1n}(\overline{\mathbf{x}}) \\ f_{21}(\overline{\mathbf{x}}) & f_{22}(\overline{\mathbf{x}}) & \cdots & f_{2n}(\overline{\mathbf{x}}) \\ \cdots & \cdots & & \cdots \\ \cdots & \cdots & & \cdots \\ f_{n1}(\overline{\mathbf{x}}) & f_{n2}(\overline{\mathbf{x}}) & \cdots & f_{nn}(\overline{\mathbf{x}}) \end{bmatrix}.$$

In expanded form, the foregoing representation can be written as

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + \sum_{j=1}^{n} f_j(\overline{\mathbf{x}})(x_j - \overline{x}_j) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \overline{x}_i)(x_j - \overline{x}_j) f_{ij}(\overline{\mathbf{x}})$$
$$+ \|\mathbf{x} - \overline{\mathbf{x}}\|^2 \alpha(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}).$$

Again, without the remainder term associated with the function α, this representation is known as a *second-order* (*Taylor series*) *approximation* at (or about) the point $\overline{\mathbf{x}}$.

3.3.6 Examples

Example 1. Let $f(x_1, x_2) = 2x_1 + 6x_2 - 2x_1^2 - 3x_2^2 + 4x_1x_2$. Then we have

$$\nabla f(\overline{\mathbf{x}}) = \begin{bmatrix} 2 - 4\overline{x}_1 + 4\overline{x}_2 \\ 6 - 6\overline{x}_2 + 4\overline{x}_1 \end{bmatrix} \quad \text{and} \quad \mathbf{H}(\overline{\mathbf{x}}) = \begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}.$$

For example, taking $\overline{\mathbf{x}} = (0,0)^t$, the second-order expansion of this function is given by

$$f(x_1, x_2) = (2,6)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \frac{1}{2}(x_1, x_2)\begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Note that there is no remainder term here since the given function is quadratic, so the above representation is exact.

Example 2. Let $f(x_1, x_2) = e^{2x_1 + 3x_2}$. Then we get

$$\nabla f(\overline{\mathbf{x}}) = \begin{bmatrix} 2e^{2\overline{x}_1 + 3\overline{x}_2} \\ 3e^{2\overline{x}_1 + 3\overline{x}_2} \end{bmatrix} \quad \text{and} \quad \mathbf{H}(\overline{\mathbf{x}}) = \begin{bmatrix} 4e^{2\overline{x}_1 + 3\overline{x}_2} & 6e^{2\overline{x}_1 + 3\overline{x}_2} \\ 6e^{2\overline{x}_1 + 3\overline{x}_2} & 9e^{2\overline{x}_1 + 3\overline{x}_2} \end{bmatrix}.$$

Hence, the second-order expansion of this function about the point $\overline{\mathbf{x}} = (2,1)^t$ is given by

$$f(\overline{\mathbf{x}}) = e^7 + (2e^7, 3e^7)\begin{pmatrix} x_1 - 2 \\ x_2 - 1 \end{pmatrix} + \frac{1}{2}(x_1 - 2, x_2 - 1)\begin{bmatrix} 4e^7 & 6e^7 \\ 6e^7 & 9e^7 \end{bmatrix}\begin{pmatrix} x_1 - 2 \\ x_2 - 1 \end{pmatrix}$$
$$+ \|\mathbf{x} - \overline{\mathbf{x}}\|^2 \alpha(\overline{\mathbf{x}}; \mathbf{x} - \overline{\mathbf{x}}).$$

Theorem 3.3.7 shows that f is convex on S if and only if its Hessian matrix is *positive semidefinite* (PSD) everywhere in S; that is, for any $\overline{\mathbf{x}}$ in S, we have $\mathbf{x}^t\mathbf{H}(\overline{\mathbf{x}})\mathbf{x} \geq 0$ for all $\mathbf{x} \in R^n$. Symmetrically, a function f is concave on S if and only if its Hessian matrix is *negative semidefinite* (NSD) everywhere in S, that is, for any $\bar{\mathbf{x}} \in S$, we have $\mathbf{x}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} \le 0$ for all $\mathbf{x} \in R^n$. A matrix that is neither positive nor negative semidefinite is called *indefinite* (ID).

3.3.7 Theorem

Let S be a nonempty open convex set in R^n, and let f: $S \to R$ be twice differentiable on S. Then f is convex if and only if the Hessian matrix is positive semidefinite at each point in S.

Proof

Suppose that f is convex, and let $\bar{\mathbf{x}} \in S$. We need to show that $\mathbf{x}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} \ge 0$ for each $\mathbf{x} \in R^n$. Since S is open, then for any given $\mathbf{x} \in R^n$, $\bar{\mathbf{x}} + \lambda\mathbf{x} \in S$ for $|\lambda| \ne 0$ and sufficiently small. By Theorem 3.3.3 and by the twice differentiability of f, we get the following two expressions:

$$f(\bar{\mathbf{x}} + \lambda\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{x} \tag{3.12}$$

$$f(\bar{\mathbf{x}} + \lambda\mathbf{x}) = f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{x} + \frac{1}{2}\lambda^2\mathbf{x}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} + \lambda^2\|\mathbf{x}\|^2\alpha(\bar{\mathbf{x}};\lambda\mathbf{x}). \tag{3.13}$$

Subtracting (3.13) from (3.12), we get

$$\frac{1}{2}\lambda^2\mathbf{x}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} + \lambda^2\|\mathbf{x}\|^2\alpha(\bar{\mathbf{x}};\lambda\mathbf{x}) \ge 0.$$

Dividing by $\lambda^2 > 0$ and letting $\lambda \to 0$, it follows that $\mathbf{x}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} \ge 0$. Conversely, suppose that the Hessian matrix is positive semidefinite at each point in S. Consider $\mathbf{x}$ and $\bar{\mathbf{x}}$ in S. Then, by the mean value theorem, we have

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^t\mathbf{H}(\hat{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}), \tag{3.14}$$

where $\hat{\mathbf{x}} = \lambda\bar{\mathbf{x}} + (1-\lambda)\mathbf{x}$ for some $\lambda \in (0, 1)$. Note that $\hat{\mathbf{x}} \in S$ and hence, by assumption, $\mathbf{H}(\hat{\mathbf{x}})$ is positive semidefinite. Therefore, $(\mathbf{x} - \bar{\mathbf{x}})^t\mathbf{H}(\hat{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \ge 0$, and from (3.14), we conclude that

$$f(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}).$$

Since the above inequality is true for each $\mathbf{x}$, $\bar{\mathbf{x}}$ in S, f is convex by Theorem 3.3.3. This completes the proof.

Theorem 3.3.7 is useful in checking the convexity or concavity of a twice differentiable function. In particular, if the function is quadratic, the Hessian matrix is independent of the point under consideration. Hence, checking its convexity reduces to checking the positive semidefiniteness of a constant matrix.

Results analogous to Theorem 3.3.7 can be obtained for the strict convex and concave cases. It turns out that if the Hessian matrix is positive definite at each point in S, the function is strictly convex. In other words, if for any given point $\overline{\mathbf{x}}$ in S, we have $\mathbf{x}^t\mathbf{H}(\overline{\mathbf{x}})\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ in R^n, then f is strictly convex. This follows readily from the proof of Theorem 3.3.7. However, if f is strictly convex, its Hessian matrix is positive semidefinite, but not necessarily positive definite everywhere in S, unless, for example, if f is quadratic. The latter is seen by writing (3.12) as a strict inequality for $\lambda\mathbf{x} \neq \mathbf{0}$ and noting that the remainder term in (3.13) is then absent. To illustrate, consider the strictly convex function defined by $f(x) = x^4$.The Hessian matrix $\mathrm{H}(x) = 12x^2$ is positive definite for all nonzero x but is positive semidefinite, and not positive definite, at $x = 0$. The following theorem records this fact.

3.3.8 Theorem

Let S be a nonempty open convex set in R^n, and let f: $S \to R$ be twice differentiable on S. If the Hessian matrix is positive definite at each point in S, f is strictly convex. Conversely, if f is strictly convex, the Hessian matrix is positive semidefinite at each point in S. However, if f is strictly convex and quadratic, its Hessian is positive definite.

The foregoing result can be strengthened somewhat while providing some additional insights into the second-order characterization of convexity. Consider, for example, the univariate function $f(x) = x^4$ addressed above, and let us show how we can argue that this function is strictly convex despite the fact that $f''(0) = 0$. Since $f''(x) \geq 0$ for all $x \in R$, we have by Theorem 3.3.7 that f is convex. Hence, by Theorem 3.3.3, all that we need to show is that for any point $\overline{x}$, the supporting hyperplane $y = f(\overline{x}) + f'(\overline{x})(x - \overline{x})$ to the epigraph of the function touches this epigraph only at the given point $(x, y) = (\overline{x}, f(\overline{x}))$. On the contrary, if this supporting hyperplane also touches the epigraph at some other point $(\hat{x}, f(\hat{x}))$, we have $f(\hat{x}) = f(\overline{x}) + f'(\overline{x})(\hat{x} - \overline{x})$. But this means that for any $x_\lambda = \lambda\overline{x} + (1-\lambda)\hat{x}$, $0 \leq \lambda \leq 1$, we have, upon using Theorem 3.3.3 and the convexity of f,

$$\lambda f(\overline{x}) + (1-\lambda)f(\hat{x}) = f(\overline{x}) + f'(\overline{x})(x_\lambda - \overline{x}) \leq f(x_\lambda) \leq \lambda f(\overline{x}) + (1-\lambda)f(\hat{x}).$$

Hence, equality holds true throughout, and the supporting hyperplane touches the graph of the function at all convex combinations $(x_\lambda, f(x_\lambda))$ as well. In fact, we obtain $f(x_\lambda) = \lambda f(\overline{x}) + (1-\lambda)f(\hat{x})$ for all $0 \leq \lambda \leq 1$, so $f''(x_\lambda) = 0$ at the uncountably infinite number of points x_λ for all $0 < \lambda < 1$. This contradicts the fact that $f''(x) = 0$ only at $x = 0$ from the above example, and therefore, the function is strictly convex. As a result, if we lose positive definiteness of a univariate convex function at only a finite (or countably infinite) number of points, we can still claim that this function is strictly convex.

Staying with univariate functions for the time being, if the function is infinitely differentiable, we can derive a necessary and sufficient condition for the function to be strictly convex. [By an *infinitely differentiable function f*: $R^n \to R$, we mean one for which for any $\overline{\mathbf{x}}$ in R^n, derivatives of all orders exist and so are continuous; are uniformly bounded in values; and for which the infinite Taylor series expansion of $f(\mathbf{x})$ about $f(\overline{\mathbf{x}})$ gives an infinite series representation of the value of f. Of course, this infinite series can possibly have only a finite number of terms, as, for example, when derivatives of order exceeding some value all vanish.]

3.3.9 Theorem

Let S be a nonempty open convex set in R, and let f: $S \to R$ be infinitely differentiable. Then f is strictly convex on S if and only if for each $\overline{x} \in S$, there exists an even n such that $f^{(n)}(\overline{x}) > 0$, while $f^{(j)}(\overline{x}) = 0$ for any $1 < j < n$, where $f^{(j)}$ denotes the jth-order derivative of f.

Proof

Let $\overline{x}$ be any point in S, and consider the infinite Taylor series expansion of f about $\overline{x}$ for a perturbation $h \neq 0$ and small enough:

$$f(\overline{x}+h) = f(\overline{x}) + hf'(\overline{x}) + \frac{h^2}{2!}f''(\overline{x}) + \frac{h^3}{3!}f'''(\overline{x}) + \cdots$$

If f is strictly convex, then by Theorem 3.3.3 we have that $f(\overline{x}+h) > f(\overline{x}) + hf'(\overline{x})$ for $h \neq 0$. Using this above, we get that for all $h \neq 0$ and sufficiently small,

$$\frac{h^2}{2!}f''(\overline{x}) + \frac{h^3}{3!}f'''(\overline{x}) + \frac{h^4}{4!}f^{(4)}(\overline{x}) + \cdots > 0.$$

Hence, not all derivatives of order greater than or equal to 2 at $\overline{x}$ can be zero. Moreover, since by making h sufficiently small, we can make the first nonzero term above dominate the rest of the expansion, and since h can be of either sign, it follows that this first nonzero derivative must be of an even order and positive for the inequality to hold true.

Conversely, suppose that given any $\overline{x} \in S$, there exists an even n such that $f^{(n)}(\overline{x}) > 0$, while $f^{(j)}(\overline{x}) = 0$ for $1 < j < n$. Then, as above, we have $(\overline{x}+h) \in S$ and $f(\overline{x}+h) > f(\overline{x}) + hf'(\overline{x})$ for all $-\delta < h < \delta$, for some $\delta > 0$ and sufficiently small. Now the hypothesis given also asserts that $f''(\overline{x}) \geq 0$ for all $\overline{x} \in S$, so by Theorem 3.3.7 we know that f is convex. Consequently, for any $\overline{h} \neq 0$, with $(\overline{x}+\overline{h}) \in S$, we get $f(\overline{x}+\overline{h}) \geq f(\overline{x}) + \overline{h}f'(\overline{x})$ by Theorem 3.3.3. To complete the proof, we must show that this inequality is indeed strict. On the contrary, if $f(\bar{x}+\bar{h}) = f(\bar{x}) + \bar{h}f'(\bar{x})$, we get

$$\begin{aligned}\lambda f(\bar{x}+\bar{h}) + (1-\lambda)f(\bar{x}) &= f(\bar{x}) + \lambda\bar{h}f'(\bar{x}) \le f(\bar{x}+\lambda\bar{h}) \\ &= f[\lambda(\bar{x}+\bar{h}) + (1-\lambda)\bar{x}] \le \lambda f(\bar{x}+\bar{h}) + (1-\lambda)f(\bar{x})\end{aligned}$$

for all $0 \le \lambda \le 1$. But this means that equality holds throughout and that $f(\bar{x} + \lambda\bar{h}) = f(\bar{x}) + \lambda\bar{h}f'(\bar{x})$ for all $0 \le \lambda \le 1$. By taking λ close enough to zero, we can contradict the statement that $f(\bar{x}+h) > f(\bar{x}) + hf'(\bar{x})$ for all $-\delta < h < \delta$, and this completes the proof.

To illustrate, when $f(x) = x^4$, we have $f'(x) = 4x^3$ and $f''(x) = 12x^2$. Hence, for $\bar{x} \ne 0$, the first nonzero derivative as in Theorem 3.3.9 is of order 2 and is positive. Furthermore, for $\bar{x} = 0$, we have $f''(\bar{x}) = f'''(\bar{x}) = 0$ and $f^{(4)}(\bar{x}) = 24 > 0$; so by Theorem 3.3.9, we can conclude that f is strictly convex.

Now let us turn to the multivariate case. The following result provides an insightful connection between the univariate and multivariate cases and permits us to derive results for the latter case from those for the former case. For notational simplicity, we have stated this result for f: $R^n \to R$, although one can readily restate it for f: $S \to R$, where S is some nonempty convex subset of R^n.

3.3.10 Theorem

Consider a function f: $R^n \to R$, and for any point $\bar{\mathbf{x}} \in R^n$ and a nonzero direction $\mathbf{d} \in R^n$, define $F_{(\bar{\mathbf{x}},\mathbf{d})}(\lambda) = f(\bar{\mathbf{x}} + \lambda\mathbf{d})$ as a function of $\lambda \in R$. Then f is (strictly) convex if and only if $F_{(\bar{\mathbf{x}};\mathbf{d})}$ is (strictly) convex for all $\bar{\mathbf{x}}$ and $\mathbf{d} \ne \mathbf{0}$ in R^n.

Proof

Given any $\bar{\mathbf{x}}$ and $\mathbf{d} \ne \mathbf{0}$ in R^n, let us write $F_{(\bar{\mathbf{x}};\mathbf{d})}(\lambda)$ simply as $F(\lambda)$ for convenience. If f is convex, then for any λ_1 and λ_2 in R and for any $0 \le \alpha \le 1$, we have

$$\begin{aligned}F(\alpha\lambda_1 + (1-\alpha)\lambda_2) &= f(\alpha[\bar{\mathbf{x}} + \lambda_1\mathbf{d}] + (1-\alpha)[\bar{\mathbf{x}} + \lambda_2\mathbf{d}]) \\ &\le \alpha f(\bar{\mathbf{x}} + \lambda_1\mathbf{d}) + (1-\alpha)f(\bar{\mathbf{x}} + \lambda_2\mathbf{d}) = \alpha F(\lambda_1) + (1-\alpha)F(\lambda_2).\end{aligned}$$

Hence, F is convex. Conversely, suppose that $F_{(\bar{\mathbf{x}};\mathbf{d})}(\lambda)$, $\lambda \in R$, is convex for all $\bar{\mathbf{x}}$ and $\mathbf{d} \ne \mathbf{0}$ in R^n. Then, for any $\mathbf{x}_1$ and $\mathbf{x}_2$ in R^n and $0 \le \lambda \le 1$, we have

$$\begin{aligned}\lambda f(\mathbf{x}_1)+(1-\lambda)f(\mathbf{x}_2) &= \lambda f[\mathbf{x}_1+0(\mathbf{x}_2-\mathbf{x}_1)]+(1-\lambda)f[\mathbf{x}_1+1(\mathbf{x}_2-\mathbf{x}_1)]\\ &= \lambda F_{[\mathbf{x}_1;(\mathbf{x}_2-\mathbf{x}_1)]}(0)+(1-\lambda)F_{[\mathbf{x}_1;(\mathbf{x}_2-\mathbf{x}_1)]}(1)\\ &\geq F_{[\mathbf{x}_1;(\mathbf{x}_2-\mathbf{x}_1)]}(1-\lambda)\\ &= f[\mathbf{x}_1+(1-\lambda)(\mathbf{x}_2-\mathbf{x}_1)]=f[\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2],\end{aligned}$$

so f is convex. The argument for the strictly convex case is similar, and this completes the proof.

This insight of examining f: $R^n \to R$ via its univariate cross sections $F_{(\bar{\mathbf{x}};\mathbf{d})}$ can be very useful both as a conceptual tool for viewing f and as an analytical tool for deriving various results. For example, writing $F(\lambda)$ = $F_{(\bar{\mathbf{x}};\mathbf{d})}(\lambda)$ = $f(\bar{\mathbf{x}}+\lambda\mathbf{d})$, for any given $\bar{\mathbf{x}}$ and $\mathbf{d}\neq\mathbf{0}$ in R^n, we have from the univariate Taylor series expansion (assuming infinite differentiability) that

$$F(\lambda)=F(0)+\lambda F'(0)+\frac{\lambda^2}{2!}F''(0)+\frac{\lambda^3}{3!}F'''(0)+\cdots.$$

By using the chain rule for differentiation, we obtain

$$F'(\lambda)=\nabla f(\bar{\mathbf{x}}+\lambda\mathbf{d})^t\mathbf{d}=\sum_i f_i(\bar{\mathbf{x}}+\lambda\mathbf{d})d_i$$

$$F''(\lambda)=\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}}+\lambda\mathbf{d})\mathbf{d}=\sum_i\sum_j f_{ij}(\bar{\mathbf{x}}+\lambda\mathbf{d})d_id_j$$

$$F'''(\lambda)=\sum_i\sum_j\sum_k f_{ijk}(\bar{\mathbf{x}}+\lambda\mathbf{d})d_id_jd_k,\text{ etc.}$$

Substituting above, this gives the corresponding multivariate Taylor series expansion as

$$f(\bar{\mathbf{x}}+\lambda\mathbf{d})=f(\bar{\mathbf{x}})+\lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{d}+\frac{\lambda^2}{2!}\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d}+\frac{\lambda^3}{3!}\sum_i\sum_j\sum_k f_{ijk}(\bar{\mathbf{x}})d_id_jd_k+\cdots.$$

As another example, using the second-order derivative result for characterizing the convexity of a univariate function along with Theorem 3.3.10, we can derive that f: $R^n \to R$ is convex if and only if $F''_{(\bar{\mathbf{x}};\mathbf{d})}(\lambda)\geq 0$ for all $\lambda\in R$, $\bar{\mathbf{x}}\in R^n$, and $\mathbf{d}\in R^n$. But since $\bar{\mathbf{x}}$ and $\mathbf{d}$ can be chosen arbitrarily, this is equivalent to requiring that $F''_{(\bar{\mathbf{x}};\mathbf{d})}(0)\geq 0$ for all $\bar{\mathbf{x}}$ and $\mathbf{d}$ in R^n. From above, this translates to the statement that $\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d}\geq 0$ for all $\mathbf{d}\in R^n$, for each $\bar{\mathbf{x}}\in R^n$, or that $\mathbf{H}(\bar{\mathbf{x}})$ is positive semidefinite for all $\bar{\mathbf{x}}\in R^n$, as in Theorem 3.3.7. In a similar manner, or by using the multivariate Taylor series expansion directly as in the proof of

Theorem 3.3.9, we can assert that an infinitely differentiable function f: $R^n \to R$ is strictly convex if and only if for each $\bar{\mathbf{x}}$ and $\mathbf{d} \neq \mathbf{0}$ in R^n, the first nonzero derivative term $[F^{(j)}(0)]$ of order greater than or equal to 2 in the Taylor series expansion above exists, is of even order, and is positive. We leave the details of exploring this result to the reader in Exercise 3.38.

We present below an efficient (polynomial-time) algorithm for checking the definiteness of a (symmetric) Hessian matrix $\mathbf{H}(\bar{\mathbf{x}})$ using elementary Gauss–Jordan operations. Appendix A cites a characterization of definiteness in terms of eigenvalues which finds use in some analytical proofs but is not an algorithmically convenient alternative. Moreover, if one needs to check for the definiteness of a matrix $\mathbf{H}(\mathbf{x})$ that is a function of $\mathbf{x}$, this eigenvalue method is very cumbersome, if not virtually impossible, to use. Although the method presented below can also get messy in such instances, it is overall a more simple and efficient approach.

We begin by considering a 2 × 2 Hessian matrix $\mathbf{H}$ in Lemma 3.3.11, where the argument $\bar{\mathbf{x}}$ has been suppressed for convenience. This is then generalized in an inductive fashion to an $n \times n$ matrix in Theorem 3.3.12.

3.3.11 Lemma

Consider a symmetric matrix $\mathbf{H} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Then $\mathbf{H}$ is positive semidefinite if and only if $a \geq 0$, $c \geq 0$, and $ac - b^2 \geq 0$, and is positive definite if and only if the foregoing inequalities are all strict.

Proof

By definition, $\mathbf{H}$ is positive semidefinite if and only if $\mathbf{d}^t\mathbf{H}\mathbf{d} = ad_1^2 + 2bd_1d_2 + cd_2^2 \geq 0$ for all $(d_1, d_2)^t \in R^2$. Hence, if $\mathbf{H}$ is positive semidefinite, we must clearly have $a \geq 0$ and $c \geq 0$. Moreover, if $a = 0$, we must have $b = 0$, so $ac - b^2 = 0$; or else, by taking $d_2 = 1$ and $d_1 = -Mb$ for $M > 0$ and large enough, we would obtain $\mathbf{d}^t\mathbf{H}\mathbf{d} < 0$, a contradiction. On the other hand, if $a > 0$, then completing the squares, we get

$$\mathbf{d}^t\mathbf{H}\mathbf{d} = a\left(d_1^2 + \frac{2bd_1d_2}{a} + \frac{b^2}{a^2}d_2^2\right) + d_2^2\left(c - \frac{b^2}{a}\right) = a\left(d_1 + \frac{b}{a}d_2\right)^2 + d_2^2\left(\frac{ac - b^2}{a}\right).$$

Hence, we must again have $(ac - b^2) \geq 0$, since otherwise, by taking $d_2 = 1$ and $d_1 = -b/a$, we would get $\mathbf{d}^t\mathbf{H}\mathbf{d} = (ac - b^2)/a < 0$, a contradiction. Hence, the condition of the theorem holds true. Conversely, suppose that $a \geq 0$, $c \geq 0$, and

$ac - b^2 \geq 0$. If $a = 0$, this gives $b = 0$, so $\mathbf{d}^t\mathbf{Hd} = cd_2^2 \geq 0$. On the other hand, if $a > 0$, by completing the squares as above we get

$$\mathbf{d}^t\mathbf{Hd} = a\left(d_1 + \frac{b}{a}d_2^2\right) + d_2^2\left(\frac{ac - b^2}{a}\right) \geq 0.$$

Hence, **H** is positive semidefinite. The proof of positive definiteness is similar, and this completes the proof.

We remark here that since a matrix **H** is negative semidefinite (negative definite) if and only if $-\mathbf{H}$ is positive semidefinite (positive definite), we get from Lemma 3.3.11 that **H** is negative semidefinite if and only if $a \leq 0$, $c \leq 0$, and $ac - b^2 \geq 0$, and that **H** is negative definite if and only if these inequalities are all strict. Theorem 3.3.12 is stated for checking positive semidefiniteness or positive definiteness of **H**. By replacing **H** by $-\mathbf{H}$, we could test symmetrically for negative semidefiniteness or negative definiteness. If the matrix turns out to be neither positive semidefinite nor negative semidefinite, it is indefinite. Also, we assume below that **H** is symmetric, being the Hessian of a twice differentiable function for our purposes. In general, if **H** is not symmetric, then since $\mathbf{d}^t\mathbf{Hd} = \mathbf{d}^t\mathbf{H}^t\mathbf{d} = \mathbf{d}^t[(\mathbf{H} + \mathbf{H}^t)/2]\mathbf{d}$, we can check for the definiteness of **H** by using the symmetric matrix $(\mathbf{H} + \mathbf{H}^t)/2$ below.

3.3.12 Theorem (Checking for PSD/PD)

Let **H** be a symmetric $n \times n$ matrix with elements h_{ij}.

(a) If $h_{ii} \leq 0$ for any $i \in \{1,\ldots,n\}$, **H** is not positive definite; and if $h_{ii} < 0$ for any $i \in \{1,\ldots,n\}$, **H** is not positive semidefinite.

(b) If $h_{ii} = 0$ for any $i \in \{1,\ldots,n\}$, we must have $h_{ij} = h_{ji} = 0$ for all $j = 1,\ldots, n$ as well, or else **H** is not positive semidefinite.

(c) If $n = 1$, **H** is positive semidefinite (positive definite) if and only if $h_{11} \geq 0$ (> 0). Otherwise, if $n \geq 2$, let

$$\mathbf{H} = \begin{bmatrix} h_{11} & \mathbf{q}^t \\ \mathbf{q} & \mathbf{G} \end{bmatrix}$$

in partitioned form, where $\mathbf{q} = 0$ if $h_{11} = 0$, and otherwise, $h_{11} > 0$. Perform elementary Gauss–Jordan operations using the first row of **H** to reduce it to the following matrix in either case:

$$\mathbf{H} = \begin{bmatrix} h_{11} & \mathbf{q}^t \\ \mathbf{0} & \mathbf{G}_{\text{new}} \end{bmatrix}.$$

Then $\mathbf{G}_{\text{new}}$ is a symmetric $(n-1) \times (n-1)$ matrix, and $\mathbf{H}$ is positive semidefinite if and only if $\mathbf{G}_{\text{new}}$ is positive semidefinite. Moreover, if $h_{11} > 0$, $\mathbf{H}$ is positive definite if and only if $\mathbf{G}_{\text{new}}$ is positive definite.

Proof

(a) Since $\mathbf{d}^t\mathbf{H}\mathbf{d} = d_i^2 h_{ii}$ whenever $d_j = 0$ for all $j \neq i$, Part (a) of the theorem is obviously true.

(b) Suppose that for some $i \neq j$, we have $h_{ii} = 0$ and $h_{ij} \neq 0$. Then, by taking $d_k = 0$ for all $k \neq i$ or j, we get $\mathbf{d}^t\mathbf{H}\mathbf{d} = 2h_{ij}d_i d_j + d_j^2 h_{jj}$, which can be made negative as in the proof of Lemma 3.3.11 by taking $d_j = 1$ and $d_i = -h_{ij}M$ for $M > 0$ and sufficiently large. This establishes Part (b).

(c) Finally, suppose that H is given in partitioned form as in Part (c). If $n = 1$, the result is trivial. Otherwise, for $n \geq 2$, let $\mathbf{d}^t = (d_1, \boldsymbol{\delta}^t)$. If $h_{11} = 0$, by assumption we also have $\mathrm{q} = 0$, and then $\mathbf{G}_{\text{new}} = \mathbf{G}$. Moreover, in this case, $\mathbf{d}^t\mathbf{H}\mathbf{d} = \boldsymbol{\delta}^t\mathbf{G}_{\text{new}}\boldsymbol{\delta}$, so $\mathbf{H}$ is positive semidefinite if and only if $\mathbf{G}_{\text{new}}$ is positive semidefinite. On the other hand, if $h_{11} > 0$, we get

$$\mathbf{d}^t\mathbf{H}\mathbf{d} = (d_1, \boldsymbol{\delta}^t)\begin{bmatrix} h_{11} & \mathrm{q}^t \\ \mathrm{q} & \mathbf{G} \end{bmatrix}\begin{pmatrix} d_1 \\ \boldsymbol{\delta} \end{pmatrix} = d_1^2 h_{11} + 2d_1(\mathrm{q}^t\boldsymbol{\delta}) + \boldsymbol{\delta}^t\mathbf{G}\boldsymbol{\delta}.$$

But by the Gauss–Jordan reduction process, we have

$$\mathbf{G}_{\text{new}} = \mathbf{G} - \frac{1}{h_{11}}\begin{bmatrix} q_1\mathrm{q}^t \\ q_2\mathrm{q}^t \\ \vdots \\ q_n\mathrm{q}^t \end{bmatrix} = \mathbf{G} - \frac{1}{h_{11}}\mathrm{q}\mathrm{q}^t,$$

which is a symmetric matrix. By substituting this above, we get

$$\mathbf{d}^t\mathbf{H}\mathbf{d} = d_1^2 h_{11} + 2d_1(\mathrm{q}^t\boldsymbol{\delta}) + \boldsymbol{\delta}^t\left(\mathbf{G}_{\text{new}} + \frac{1}{h_{11}}\mathrm{q}\mathrm{q}^t\right)\boldsymbol{\delta} = \boldsymbol{\delta}^t\mathbf{G}_{\text{new}}\boldsymbol{\delta} + h_{11}\left(d_1 + \frac{\mathrm{q}^t\boldsymbol{\delta}}{h_{11}}\right)^2.$$

Hence, it can readily be verified that $\mathbf{d}^t\mathbf{H}\mathbf{d} \geq 0$ for all $\mathbf{d} \in R^n$ if and only if $\boldsymbol{\delta}^t\mathbf{G}_{\text{new}}\boldsymbol{\delta} \geq 0$ for all $\boldsymbol{\delta} \in R^{n-1}$, because $h_{11}(d_1 + \mathrm{q}^t\boldsymbol{\delta}/h_{11})^2 \geq 0$, and the latter term can be made zero by selecting $d_1 = -\mathrm{q}^t\boldsymbol{\delta}/h_{11}$, if necessary. By the same argument, $\mathbf{d}^t\mathbf{H}\mathbf{d} > 0$ for all $\mathbf{d} \neq \mathbf{0}$ in R^n if and only if $\delta^t\mathbf{G}_{\text{new}}\delta > 0$ for all $\delta \neq \mathbf{0}$ in R^{n-1}, and this completes the proof.

Observe that Theorem 3.3.12 prompts a polynomial-time algorithm for checking the PSD/PD of a symmetric $n \times n$ matrix **H**. We first scan the diagonal elements to see if either condition (a) or (b) leads to the conclusion that the matrix is not PSD/PD. If this does not terminate the process, we perform the Gauss–Jordan reduction as in Part (c) and arrive at a matrix $\mathbf{G}_{\text{new}}$ of one lesser dimension for which we may now perform the same test as on **H**. When $\mathbf{G}_{\text{new}}$ is finally a 2×2 matrix, we can use Lemma 3.3.11, or we can continue to reduce it to a 1×1 matrix and hence determine the PSD/PD of **H**. Since each pass through the inductive step of the algorithm is of *complexity* $O(n^2)$ (read as "*of order* n^2" and meaning that the number of elementary arithmetic operations, comparison, etc., involved are bounded above by Kn^2 for some constant K) and the number of inductive steps is of $O(n)$, the overall process is of polynomial complexity $O(n^3)$. Because the algorithm basically works toward reducing the matrix to an upper triangular matrix, it is sometimes called a *superdiagonalization algorithm*. This algorithm affords a proof for the following useful result, which can alternatively be proved using the eigenvalue characterization of definiteness (see Exercise 3.42).

Corollary

Let **H** be an $n \times n$ symmetric matrix. Then **H** is positive definite if and only if it is positive semidefinite and nonsingular.

Proof

If **H** is positive definite, it is positive semidefinite; and since the superdiagonalization algorithm reduces the matrix **H** to an upper triangular matrix with positive diagonal elements via elementary row operations, **H** is nonsingular. Conversely, if **H** is positive semidefinite and nonsingular, the superdiagonalization algorithm must always encounter nonzero elements along the diagonal because **H** is nonsingular, and these must be positive because **H** is positive semidefinite. Hence, **H** is positive definite.

3.3.13 Examples

Example 1. Consider Example 1 of Section 3.3.6. Here we have

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}$$

so

$$-\mathbf{H}(\mathbf{x}) = \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix}.$$

By Lemma 3.3.11 we conclude that $-\mathbf{H}(\mathbf{x})$ is positive definite, so $\mathbf{H}(\mathbf{x})$ is negative definite and the function f is strictly concave.

Example 2. Consider the function $f(x_1, x_2) = x_1^3 + 2x_2^2$. Here we have

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 3x_1^2 \\ 4x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{H}(\mathbf{x}) = \begin{bmatrix} 6x_1 & 0 \\ 0 & 4 \end{bmatrix}.$$

By Lemma 3.3.11, whenever $x_1 < 0$, $\mathbf{H}(\mathbf{x})$ is indefinite. However, $\mathbf{H}(\mathbf{x})$ is positive definite for $x_1 > 0$, so f is strictly convex over $\{\mathbf{x} : x_1 > 0\}$.

Example 3. Consider the matrix

$$\mathbf{H} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}.$$

Note that the matrix is not negative semidefinite. To check PSD/PD, apply the superdiagonalization algorithm and reduce $\mathbf{H}$ to

$$\begin{bmatrix} 2 & 1 & 2 \\ 0 & 3/2 & 2 \\ 0 & 2 & 2 \end{bmatrix} \quad \text{which gives} \quad \mathbf{G}_{\text{new}} = \begin{bmatrix} 3/2 & 2 \\ 2 & 2 \end{bmatrix}.$$

Now the diagonals of $\mathbf{G}_{\text{new}}$ are positive, but $\det(\mathbf{G}_{\text{new}}) = -1$. Hence, $\mathbf{H}$ is not positive semidefinite. Alternatively, we could have verified this by continuing to reduce $\mathbf{G}_{\text{new}}$ to obtain the matrix

$$\begin{bmatrix} 3/2 & 2 \\ 0 & -2/3 \end{bmatrix}.$$

Since the resulting second diagonal element (i.e., the reduced $\mathbf{G}_{\text{new}}$) is negative, $\mathbf{H}$ is not positive semidefinite. Since $\mathbf{H}$ is not negative semidefinite either, it is indefinite.

3.4 Minima and Maxima of Convex Functions

In this section we consider the problems of minimizing and maximizing a convex function over a convex set and develop necessary and/or sufficient conditions for optimality.

Minimizing a Convex Function

The case of maximizing a concave function is similar to that of minimizing a convex function. We develop the latter in detail and ask the reader to draw the analogous results for the concave case.

3.4.1 Definition

Let f: $R^n \to R$ and consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. A point $\mathbf{x} \in S$ is called a *feasible solution* to the problem. If $\overline{\mathbf{x}} \in S$ and $f(\mathbf{x}) \geq f(\overline{\mathbf{x}})$ for each $\mathbf{x} \in S$, $\overline{\mathbf{x}}$ is called an *optimal solution*, a *global optimal solution*, or simply a *solution* to the problem. The collection of optimal solutions are called *alternative optimal solutions*. If $\overline{\mathbf{x}} \in S$ and if there exists an ε-neighborhood $N_\varepsilon(\overline{\mathbf{x}})$ around $\overline{\mathbf{x}}$ such that $f(\mathbf{x}) \geq f(\overline{\mathbf{x}})$ for each $\mathbf{x} \in S \cap N_\varepsilon(\overline{\mathbf{x}})$, $\overline{\mathbf{x}}$ is called a *local optimal solution*. Similarly, if $\overline{\mathbf{x}} \in S$ and if $f(\mathbf{x}) > f(\overline{\mathbf{x}})$ for all $\mathbf{x} \in S \cap N_\varepsilon(\overline{\mathbf{x}})$, $\mathbf{x} \neq \overline{\mathbf{x}}$, for some $\varepsilon > 0$, $\overline{\mathbf{x}}$ is called a *strict local optimal solution*. On the other hand, if $\overline{\mathbf{x}} \in S$ is the *only* local minimum in $S \cap N_\varepsilon(\overline{\mathbf{x}})$, for some ε-neighborhood $N_\varepsilon(\overline{\mathbf{x}})$ around $\overline{\mathbf{x}}$, $\overline{\mathbf{x}}$ is called a *strong* or *isolated local optimal solution*. All these types of local optima or minima are sometimes also referred to as *relative minima*. Figure 3.6 illustrates instances of local and global minima for the problem of minimizing $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f and S are shown in the figure.

The points in S corresponding to A, B, and C are also both strict and strong local minima, whereas those corresponding to the flat segment of the graph between D and E are local minima that are neither strict nor strong. Note that if $\overline{\mathbf{x}}$ is a strong or isolated local minimum, it is also a strict minimum. To see this, consider the ε-neighborhood $N_\varepsilon(\overline{\mathbf{x}})$ characterizing the strong local minimum nature of $\overline{\mathbf{x}}$. Then we must also have $f(\mathbf{x}) > f(\overline{\mathbf{x}})$ for all $\mathbf{x} \in S \cap$

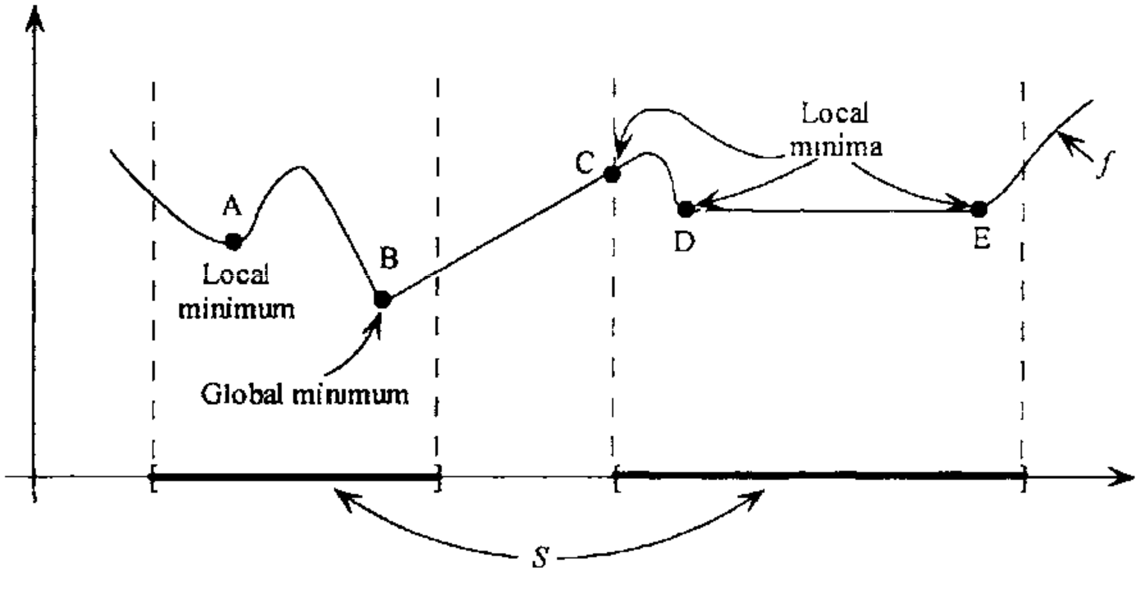


Figure 3.6 Local and global minima.

$N_\varepsilon(\bar{\mathbf{x}})$, because otherwise, suppose that there exists an $\hat{\mathbf{x}} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$ such that $f(\hat{\mathbf{x}}) = f(\bar{\mathbf{x}})$. Note that $\hat{\mathbf{x}}$ is an alternative optimal solution within $S \cap N_\varepsilon(\bar{\mathbf{x}})$, so there exists some $0 < \varepsilon' < \varepsilon$ such that $f(\mathbf{x}) \geq f(\hat{\mathbf{x}})$ for all $\mathbf{x} \in S \cap N_{\varepsilon'}(\hat{\mathbf{x}})$. But this contradicts the isolated local minimum status of $\bar{\mathbf{x}}$, and hence $\bar{\mathbf{x}}$ must also be a strict local minimum. On the other hand, a strict local minimum need not be an isolated local minimum. Figure 3.7 illustrates two such instances. In Figure 3.7*a*, $S = R$ and $f(x) = 1$ for $x = 1$ and is equal to 2 otherwise. Note that the point of discontinuity $\bar{x} = 1$ of f is a strict local minimum but is not isolated, since any ε-neighborhood about $\bar{x}$ contains points other than $\bar{x} = 1$, all of which are also local minima. Figure 3.7*b* illustrates another case in which $f(x) = x^2$, a strictly convex function; but $S = \{1/2^k, k = 0,1,2,...\} \cup \{0\}$ is a nonconvex set. Here, the point $\bar{\mathbf{x}} = 1/2^k$ for any integer $k \geq 0$ is an isolated and therefore a strict local minimum because it can be captured as the unique feasible solution in $S \cap N_\varepsilon(\bar{\mathbf{x}})$ for some sufficiently small $\varepsilon > 0$. However, although $\bar{x} = 0$ is clearly a strict local minimum (it is, in fact, the unique global minimum), it is not isolated because any ε-neighborhood about $\bar{x} = 0$ contains other local minima of the foregoing type.

Nonetheless, for optimization problems, $\min\{f(\mathbf{x}) : \mathbf{x} \in S\}$, where f is a convex function and S is a convex set, which are known as *convex programming problems* and that are of interest to us in this section, a strict local minimum is also a strong local minimum, as shown in Theorem 3.4.2 (see Exercise 3.47 for a weaker sufficient condition). The principal result here is that each local minimum of a convex program is also a global minimum. This fact is quite useful in the optimization process, since it enables us to stop with a global optimal solution if the search in the vicinity of a feasible point does not lead to an improving feasible solution.

3.4.2 Theorem

Let S be a nonempty convex set in R^n, and let f: $S \to R$ be convex on S. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Suppose that $\bar{\mathbf{x}} \in S$ is a local optimal solution to the problem.

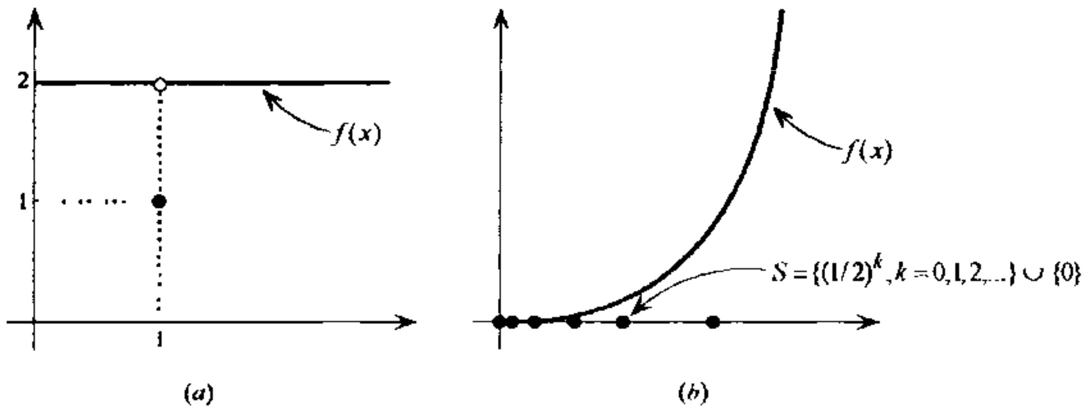


Figure 3.7 Strict local minima are not necessarily strong local minima.

1. Then $\bar{\mathbf{x}}$ is a global optimal solution.
2. If either $\bar{\mathbf{x}}$ is a strict local minimum or f is strictly convex, $\bar{\mathbf{x}}$ is the unique global optimal solution and is also a strong local minimum.

Proof

Since $\bar{\mathbf{x}}$ is a local optimal solution, there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ around $\bar{\mathbf{x}}$ such that

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) \qquad \text{for each } \mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}}). \tag{3.15}$$

By contradiction, suppose that $\bar{\mathbf{x}}$ is not a global optimal solution so that $f(\hat{\mathbf{x}}) < f(\bar{\mathbf{x}})$ for some $\hat{\mathbf{x}} \in S$. By the convexity of f, the following is true for each $0 \leq \lambda \leq 1$:

$$f(\lambda\hat{\mathbf{x}} + (1-\lambda)\bar{\mathbf{x}}) \leq \lambda f(\hat{\mathbf{x}}) + (1-\lambda) f(\bar{\mathbf{x}}) < \lambda f(\bar{\mathbf{x}}) + (1-\lambda) f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}).$$

But for $\lambda > 0$ and sufficiently small, $\lambda\hat{\mathbf{x}} + (1-\lambda)\bar{\mathbf{x}} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$. Hence, the above inequality contradicts (3.15), and Part 1 is proved.

Next, suppose that $\bar{\mathbf{x}}$ is a strict local minimum. By Part 1 it is a global minimum. Now, on the contrary, if there exists an $\hat{\mathbf{x}} \in S$ such that $f(\hat{\mathbf{x}}) = f(\bar{\mathbf{x}})$, then defining $\mathbf{x}_\lambda = \lambda\hat{\mathbf{x}} + (1-\lambda)\bar{\mathbf{x}}$ for $0 \leq \lambda \leq 1$, we have, by the convexity of f and S that $f(\mathbf{x}_\lambda) \leq \lambda f(\hat{\mathbf{x}}) + (1-\lambda) f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}})$, and $\mathbf{x}_\lambda \in S$ for all $0 \leq \lambda \leq 1$. By taking $\lambda \to 0^+$, since we can make $\mathbf{x}_\lambda \in N_\varepsilon(\bar{\mathbf{x}}) \cap S$ for any $\varepsilon > 0$, this contradicts the strict local optimality of $\bar{\mathbf{x}}$. Hence, $\bar{\mathbf{x}}$ is the unique global minimum. Therefore, it must also be an isolated local minimum, since any other local minimum in $N_\varepsilon(\bar{\mathbf{x}}) \cap S$ for any $\varepsilon > 0$ would also be a global minimum, which is a contradiction.

Finally, suppose that $\bar{\mathbf{x}}$ is a local optimal solution and that f is strictly convex. Since strict convexity implies convexity, then by Part 1, $\bar{\mathbf{x}}$ is a global optimal solution. By contradiction, suppose that $\bar{\mathbf{x}}$ is not the unique global optimal solution, so that there exists an $\mathbf{x} \in S$, $\mathbf{x} \neq \bar{\mathbf{x}}$ such that $f(\mathbf{x}) = f(\bar{\mathbf{x}})$. By strict convexity,

$$f\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\bar{\mathbf{x}}\right) < \frac{1}{2} f(\mathbf{x}) + \frac{1}{2} f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}).$$

By the convexity of S, $(1/2)\mathbf{x} + (1/2)\bar{\mathbf{x}} \in S$, and the above inequality violates the global optimality of $\bar{\mathbf{x}}$. Hence, $\bar{\mathbf{x}}$ is the unique global minimum and, as above, is also a strong local minimum. This completes the proof.

We now develop a necessary and sufficient condition for the existence of a global solution. If such an optimal solution does not exist, then $\inf\{f(\mathbf{x}) : \mathbf{x} \in S\}$ is finite but is not achieved at any point in S, or it is equal to $-\infty$.

3.4.3 Theorem

Let $f: R^n \to R$ be a convex function, and let S be a nonempty convex set in R^n. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. The point $\bar{\mathbf{x}} \in S$ is an optimal solution to this problem if and only if f has a subgradient ξ at $\bar{\mathbf{x}}$ such that $\xi^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$.

Proof

Suppose that $\xi^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$, where ξ is a subgradient of f at $\bar{\mathbf{x}}$. By the convexity of f, we have

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \xi^t(\mathbf{x}-\bar{\mathbf{x}}) \geq f(\bar{\mathbf{x}}) \qquad \text{for all } \mathbf{x} \in S,$$

and hence $\bar{\mathbf{x}}$ is an optimal solution to the given problem.

To show the converse, suppose that $\bar{\mathbf{x}}$ is an optimal solution to the problem and construct the following two sets in R^{n+1}:

$$\Lambda_1 = \{(\mathbf{x}-\bar{\mathbf{x}}, y) : \mathbf{x} \in R^n,\ y > f(\mathbf{x}) - f(\bar{\mathbf{x}})\}$$
$$\Lambda_2 = \{(\mathbf{x}-\bar{\mathbf{x}}, y) : \mathbf{x} \in S,\ y \leq 0\}.$$

The reader may easily verify that both Λ_1 and Λ_2 are convex sets. Also, $\Lambda_1 \cap \Lambda_2 = \varnothing$ because otherwise there would exist a point $(\mathbf{x}, y)$ such that

$$\mathbf{x} \in S, \qquad 0 \geq y > f(\mathbf{x}) - f(\bar{\mathbf{x}}),$$

contradicting the assumption that $\bar{\mathbf{x}}$ is an optimal solution to the problem. By Theorem 2.4.8 there is a hyperplane that separates Λ_1 and Λ_2; that is, there exist a nonzero vector (ξ_0, μ) and a scalar α such that

$$\xi_0^t(\mathbf{x}-\bar{\mathbf{x}}) + \mu y \leq \alpha, \quad \forall \mathbf{x} \in R^n,\ y > f(\mathbf{x}) - f(\bar{\mathbf{x}}) \tag{3.16}$$

$$\xi_0^t(\mathbf{x}-\bar{\mathbf{x}}) + \mu y \geq \alpha, \quad \forall \mathbf{x} \in S,\ y \leq 0. \tag{3.17}$$

If we let $\mathbf{x} = \bar{\mathbf{x}}$ and $y = 0$ in (3.17), it follows that $\alpha \leq 0$. Next, letting $\mathbf{x} = \bar{\mathbf{x}}$ and $y = \varepsilon > 0$ in (3.16), it follows that $\mu\varepsilon \leq \alpha$. Since this is true for every $\varepsilon > 0$, $\mu \leq 0$ and $\alpha \geq 0$. To summarize, we have shown that $\mu \leq 0$ and $\alpha = 0$. If $\mu = 0$, from (3.16), $\xi_0^t(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$ for each $\mathbf{x} \in R^n$. If we let $\mathbf{x} = \bar{\mathbf{x}} + \xi_0$, it follows that

$$0 \geq \xi_0^t(\mathbf{x}-\bar{\mathbf{x}}) = \|\xi_0\|^2$$

and hence $\xi_0 = 0$. Since $(\xi_0, \mu) \neq (0,0)$, we must have $\mu < 0$. Dividing (3.16) and (3.17) by $-\mu$ and denoting $-\xi_0/\mu$ by ξ, we get the following inequalities:

$$y \geq \xi^t(\mathbf{x}-\overline{\mathbf{x}}), \quad \forall \mathbf{x} \in R^n, \quad y > f(\mathbf{x}) - f(\overline{\mathbf{x}}) \tag{3.18}$$

$$\xi^t(\mathbf{x}-\overline{\mathbf{x}}) - y \geq 0, \quad \forall \mathbf{x} \in S, \quad y \leq 0. \tag{3.19}$$

By letting $y = 0$ in (3.19), we get $\xi^t(\mathbf{x}-\overline{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$. From (3.18) it is obvious that

$$f(\mathbf{x}) \geq f(\overline{\mathbf{x}}) + \xi^t(\mathbf{x}-\overline{\mathbf{x}}) \qquad \text{for all } \mathbf{x} \in R^n.$$

Therefore, ξ is a subgradient of f at $\overline{\mathbf{x}}$ with the property that $\xi^t(\mathbf{x}-\overline{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$, and the proof is complete.

Corollary 1

Under the assumptions of Theorem 3.4.3, if S is open, $\overline{\mathbf{x}}$ is an optimal solution to the problem if and only if there exists a zero subgradient of f at $\overline{\mathbf{x}}$. In particular, if $S = R^n$, $\overline{\mathbf{x}}$ is a global minimum if and only if there exists a zero subgradient of f at $\overline{\mathbf{x}}$.

Proof

By the theorem, $\overline{\mathbf{x}}$ is an optimal solution if and only if $\xi^t(\mathbf{x}-\overline{\mathbf{x}}) \geq 0$ for each $\mathbf{x} \in S$, where ξ is a subgradient of f at $\overline{\mathbf{x}}$. Since S is open, $\mathbf{x} = \overline{\mathbf{x}} - \lambda\xi \in S$ for some positive λ. Therefore, $-\lambda\|\xi\|^2 \geq 0$; that is, $\xi = \mathbf{0}$.

Corollary 2

In addition to the assumptions of the theorem, suppose that f is differentiable. Then $\overline{\mathbf{x}}$ is an optimal solution if and only if $\nabla f(\overline{\mathbf{x}})^t(\mathbf{x}-\overline{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$. Furthermore, if S is open, $\overline{\mathbf{x}}$ is an optimal solution if and only if $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$.

Note the important implications of Theorem 3.4.3. First, the theorem gives a necessary and sufficient characterization of optimal solutions. This characterization reduces to the well-known condition of vanishing derivatives if f is differentiable and S is open. Another important implication is that if we reach a nonoptimal point $\overline{\mathbf{x}}$, where $\nabla f(\overline{\mathbf{x}})^t(\mathbf{x}-\overline{\mathbf{x}}) < 0$ for some $\mathbf{x} \in S$, there is an obvious way to proceed to an improving solution. This can be achieved by moving from $\overline{\mathbf{x}}$ in the direction $\mathbf{d} = \mathbf{x} - \overline{\mathbf{x}}$. The actual size of the step can be determined by solving a *line search problem*, which is a one-dimensional minimization subproblem of the following form: Minimize $f[\overline{\mathbf{x}} + \lambda \mathrm{d}]$ subject to $\lambda \geq 0$ and $\overline{\mathbf{x}} + \lambda\mathbf{d} \in S$. This procedure is called the *method of feasible directions* and is discussed in more detail in Chapter 10.

To provide additional insights, let us dwell for awhile on Corollary 2, which addresses the differentiable case for Theorem 3.4.3. Figure 3.8 illustrates

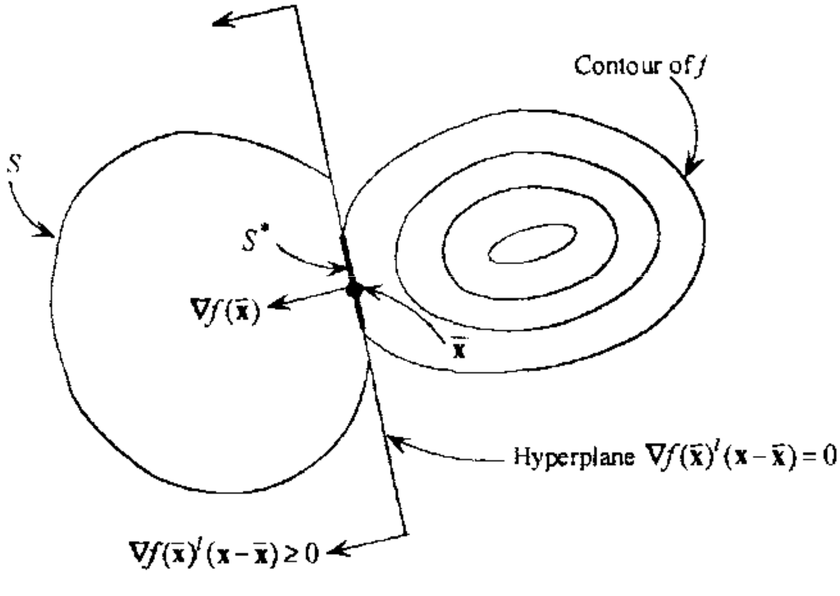


Figure 3.8 Geometry for Theorems 3.4.3 and 3.4.4.

the geometry of the result. Now suppose that for the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, we have f differentiable and convex, but S is an arbitrary set. Suppose further that it turns out that the directional derivative $f'(\bar{\mathbf{x}};\mathbf{x}-\bar{\mathbf{x}})$ = $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$. The proof of the theorem actually shows that $\bar{\mathbf{x}}$ is a global minimum regardless of S, since for any solution $\hat{\mathbf{x}}$ that improves over $\bar{\mathbf{x}}$, we have, by the convexity of f, that $f(\bar{\mathbf{x}}) > f(\hat{\mathbf{x}}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}}-\bar{\mathbf{x}})$, which implies that $\nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}}-\bar{\mathbf{x}}) < 0$, whereas $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$. Hence, the hyperplane $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) = 0$ separates S from solutions that improve over $\bar{\mathbf{x}}$. [For the nondifferentiable case, the hyperplane $\boldsymbol{\xi}^t(\mathbf{x}-\bar{\mathbf{x}}) = 0$ plays a similar role.] However, if f is not convex, the directional derivative $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})$ being nonnegative for all $\mathbf{x} \in S$ does not even necessarily imply that $\bar{\mathbf{x}}$ is a local minimum. For example, for the problem to minimize $f(x) = x^3$ subject to $-1 \leq x \leq 1$, we have the condition $f'(\bar{x})(x-\bar{x}) \geq 0$ for all $x \in S$ being satisfied at $\bar{x} = 0$, since $f'(0) = 0$, but $\bar{x} = 0$ is not even a local minimum for this problem.

Conversely, suppose that f is differentiable but arbitrary otherwise and that S is a convex set. Then, if $\bar{\mathbf{x}}$ is a global minimum, we must have $f'(\bar{\mathbf{x}};\mathbf{x}-\bar{\mathbf{x}})$ $= \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$. This follows because, otherwise, if $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) < 0$, we could move along the direction $\mathbf{d} = \mathbf{x} - \bar{\mathbf{x}}$ and, as above, the objective value would fall for sufficiently small step lengths, whereas $\bar{\mathbf{x}}+\lambda\mathbf{d}$ would remain feasible for $0 \leq \lambda \leq 1$ by the convexity of S. Note that this explains a more general concept: if f is differentiable but f and S are otherwise arbitrary, and if $\bar{\mathbf{x}}$ is a local minimum of f over S, then for any direction d for which $\bar{\mathbf{x}}+\lambda\mathbf{d}$ remains feasible for $0 < \lambda \leq \delta$ for some $\delta > 0$, we must have a nonnegative directional derivative of f at $\bar{\mathbf{x}}$ in the direction d; that is, we must have $f'(\bar{\mathbf{x}};\mathbf{d}) = \nabla f(\bar{\mathbf{x}})^t\mathbf{d} \geq 0$.

Now let us turn our attention back to convex programming problems. The following result and its corollaries characterize the set of alternative optimal solutions and show, in part, that the gradient of the objective function (assuming twice differentiability) is a constant over the optimal solution set, and that for a quadratic objective function, the optimal solution set is in fact polyhedral. (See Figure 3.8 to identify the set of alternative optimal solutions S^* defined by the theorem in light of Theorem 3.4.3.)

3.4.4 Theorem

Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f is a convex and twice differentiable function and S is a convex set, and suppose that there exists an optimal solution $\bar{\mathbf{x}}$. Then the set of alternative optimal solutions is characterized by the set

$$S^* = \{\mathbf{x} \in S : \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \leq 0 \text{ and } \nabla f(\mathbf{x}) = \nabla f(\bar{\mathbf{x}})\}.$$

Proof

Denote the set of alternative optimal solutions as $\bar{S}$, say, and note that $\bar{\mathbf{x}} \in \bar{S} \neq \varnothing$. Consider any $\hat{\mathbf{x}} \in S^*$. By the convexity of f and the definition of S^*, we have $\hat{\mathbf{x}} \in S$ and

$$f(\bar{\mathbf{x}}) \geq f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})^t(\bar{\mathbf{x}}-\hat{\mathbf{x}}) = f(\hat{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\bar{\mathbf{x}}-\hat{\mathbf{x}}) \geq f(\hat{\mathbf{x}}),$$

so we must have $\hat{\mathbf{x}} \in \bar{S}$ by the optimality of $\bar{\mathbf{x}}$. Hence, $S^* \subseteq \bar{S}$.

Conversely, suppose that $\hat{\mathbf{x}} \in \bar{S}$, so that $\hat{\mathbf{x}} \in S$ and $f(\hat{\mathbf{x}}) = f(\bar{\mathbf{x}})$. This means that $f(\bar{\mathbf{x}}) = f(\hat{\mathbf{x}}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}}-\bar{\mathbf{x}})$ or that $\nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}}-\bar{\mathbf{x}}) \leq 0$. But by Corollary 2 to Theorem 3.4.3, we have $\nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}}-\bar{\mathbf{x}}) \geq 0$. Hence, $\nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}}-\bar{\mathbf{x}}) = 0$. By interchanging the roles of $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}$, we obtain $\nabla f(\hat{\mathbf{x}})^t(\bar{\mathbf{x}}-\hat{\mathbf{x}}) = 0$ symmetrically. Therefore,

$$[\nabla f(\bar{\mathbf{x}}) - \nabla f(\hat{\mathbf{x}})]^t(\bar{\mathbf{x}}-\hat{\mathbf{x}}) = 0. \tag{3.20}$$

Now we have

$$\begin{aligned}[\nabla f(\bar{\mathbf{x}}) - \nabla f(\hat{\mathbf{x}})] &= \nabla f[\hat{\mathbf{x}} + \lambda(\bar{\mathbf{x}}-\hat{\mathbf{x}})]_{\lambda=0}^{\lambda=1} \\ &= \int_{\lambda=0}^{\lambda=1}\mathbf{H}[\hat{\mathbf{x}} + \lambda(\bar{\mathbf{x}}-\hat{\mathbf{x}})](\bar{\mathbf{x}}-\hat{\mathbf{x}})\,d\lambda = \mathbf{G}(\bar{\mathbf{x}}-\hat{\mathbf{x}}),\end{aligned} \tag{3.21}$$

where $\mathbf{G} = \int_0^1 \mathbf{H}[\hat{\mathbf{x}} + \lambda(\bar{\mathbf{x}} - \hat{\mathbf{x}})]\, d\lambda$ and where the integral of the matrix is performed componentwise. But note that $\mathbf{G}$ is positive semidefinite because $\mathbf{d}^t\mathbf{G}\mathbf{d} = \int_0^1 \mathbf{d}^t \mathbf{H}[\hat{\mathbf{x}} + \lambda(\bar{\mathbf{x}} - \hat{\mathbf{x}})]\mathbf{d}\ \ d\lambda \geq 0$ for all $\mathbf{d} \in R^n$, since $\mathbf{d}^t\mathbf{H}[\hat{\mathbf{x}} + \lambda(\bar{\mathbf{x}} - \hat{\mathbf{x}})]\mathbf{d}$ is a nonnegative function of λ by the convexity of f. Hence, by (3.20) and (3.21), we get $0 = (\bar{\mathbf{x}} - \hat{\mathbf{x}})^t[\nabla f(\bar{\mathbf{x}}) - \nabla f(\hat{\mathbf{x}})] = (\bar{\mathbf{x}} - \hat{\mathbf{x}})^t \mathbf{G}(\bar{\mathbf{x}} - \hat{\mathbf{x}})$. But the positive semidefiniteness of $\mathbf{G}$ implies that $\mathbf{G}(\bar{\mathbf{x}} - \hat{\mathbf{x}}) = \mathbf{0}$ by a standard result (see Exercise 3.41). Therefore, by (3.21), we have $\nabla f(\bar{\mathbf{x}}) = \nabla f(\hat{\mathbf{x}})$. We have hence shown that $\hat{\mathbf{x}} \in S$, $\nabla f(\bar{\mathbf{x}})^t\ (\hat{\mathbf{x}} - \bar{\mathbf{x}}) \leq 0$, and $\nabla f(\hat{\mathbf{x}}) = \nabla f(\bar{\mathbf{x}})$. This means that $\hat{\mathbf{x}} \in S^*$, and thus $\bar{S} \subseteq S^*$. This, together with $S^* \subseteq \bar{S}$, completes the proof.

Corollary 1

The set S^* of alternative optimal solutions can equivalently be defined as

$$S^* = \{\mathbf{x} \in S : \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) = 0 \text{ and } \nabla f(\mathbf{x}) = \nabla f(\bar{\mathbf{x}})\}.$$

Proof

The proof follows from the definition of S^* in Theorem 3.4.4 and the fact that $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in S$ by Corollary 2 to Theorem 3.4.3.

Corollary 2

Suppose that f is a quadratic function given by $f(\mathbf{x}) = \mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ and that S is polyhedral. Then S^* is a polyhedral set given by

$$S^* = \{\mathbf{x} \in S : \mathbf{c}^t(\mathbf{x} - \bar{\mathbf{x}}) \leq 0,\ \ \mathbf{H}(\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{0}\} = \{\mathbf{x} \in S : \mathbf{c}^t(\mathbf{x} - \bar{\mathbf{x}}) = 0,$$
$$\mathbf{H}(\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{0}\}.$$

Proof

The proof follows by direct substitution in Theorem 3.4.4 and Corollary 1, noting that $\nabla f(\mathbf{x}) = \mathbf{c} + \mathbf{H}\mathbf{x}$.

3.4.5 Example

$$\text{Minimize } \left(x_1 - \frac{3}{2}\right)^2 + (x_2 - 5)^2$$
$$\text{subject to } \begin{aligned} -x_1 + x_2 &\le 2 \\ 2x_1 + 3x_2 &\le 11 \\ -x_1 &\le 0 \\ -x_2 &\le 0. \end{aligned}$$

Clearly, $f(x_1, x_2) = (x_1 - 3/2)^2 + (x_2 - 5)^2$ is a convex function, which gives the square of the distance from the point (3/2, 5). The convex polyhedral set S is represented by the above four inequalities. The problem is depicted in Figure 3.9. From the figure, clearly the optimal point is (1, 3). The gradient vector of f at the point (1, 3) is $\nabla f(1,3) = (-1, -4)^t$. We see geometrically that the vector (–1, –4) makes an angle of $< 90°$ with each vector of the form $(x_1 - 1, x_2 - 3)$, where $(x_1, x_2) \in S$. Thus, the optimality condition of Theorem 3.4.3 is verified and, by Theorem 3.4.4, (1, 3) is the unique optimum.

To illustrate further, suppose that it is claimed that $\hat{\mathbf{x}} = (0,0)^t$ is an optimal point. By Theorem 3.4.4, this cannot be true since we have $\nabla f(\bar{\mathbf{x}})^t(\hat{\mathbf{x}} - \bar{\mathbf{x}}) = 13 > 0$ when $\bar{\mathbf{x}} = (1,3)^t$. Similarly, by Theorem 3.4.3, we can easily verify that $\hat{\mathbf{x}}$ is not optimal. Note that $\nabla f(0,0) = (-3, -10)^t$ and actually, for each nonzero $\mathbf{x} \in S$, we have $-3x_1 - 10x_2 < 0$. Hence, the origin could not be an optimal point. Moreover, we can improve f by moving from $\mathbf{0}$ in the direction $\mathbf{x} - \mathbf{0}$ for any $\mathbf{x} \in S$. In this case, the best local direction is $-\nabla f(0,0)$, that is, the direction (3, 10). In Chapter 10 we discuss methods for finding a particular direction among many alternatives.

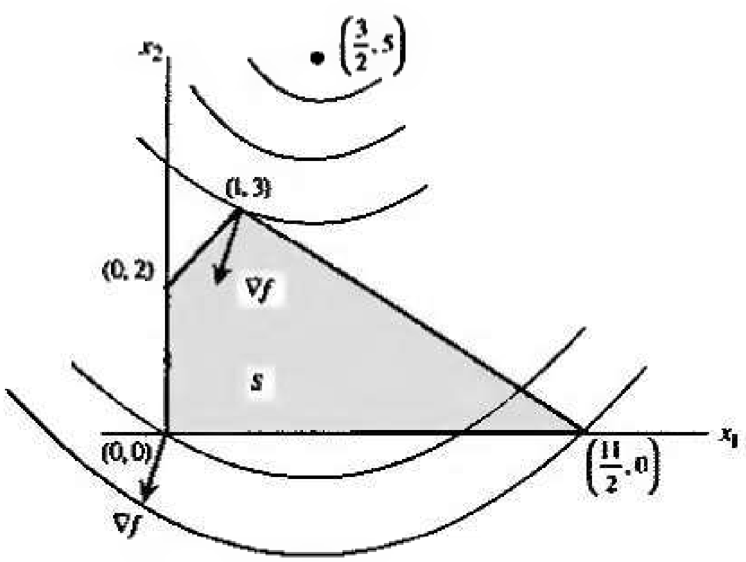


Figure 3.9 Setup for Example 3.4.5.

Maximizing a Convex Function

We now develop a necessary condition for a maximum of a convex function over a convex set. Unfortunately, this condition is not sufficient. Therefore, it is possible, and actually not unlikely, that several local maxima satisfying the condition of Theorem 3.4.6 exist. Unlike the minimization case, there exists no local information at such solutions that could lead us to better points. Hence, maximizing a convex function is usually a much harder task than minimizing a convex function. Again, minimizing a concave function is similar to maximizing a convex function, and hence the development for the concave case is left to the reader.

3.4.6 Theorem

Let f: $R^n \to R$ be a convex function, and let S be a nonempty convex set in R^n. Consider the problem to maximize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. If $\bar{\mathbf{x}} \in S$ is a local optimal solution, $\boldsymbol{\xi}^t(\mathbf{x}-\bar{\mathbf{x}}) \le 0$ for each $\mathbf{x} \in S$, where $\boldsymbol{\xi}$ is any subgradient of f at $\bar{\mathbf{x}}$.

Proof

Suppose that $\bar{\mathbf{x}} \in S$ is a local optimal solution. Then an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ exists such that $f(\mathbf{x}) \le f(\bar{\mathbf{x}})$ for each $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$. Let $\mathbf{x} \in S$, and note that $\bar{\mathbf{x}} + \lambda(\mathbf{x}-\bar{\mathbf{x}}) \in S \cap N_\varepsilon(\bar{\mathbf{x}})$ for $\lambda > 0$ and sufficiently small. Hence,

$$f[\bar{\mathbf{x}} + \lambda(\mathbf{x}-\bar{\mathbf{x}})] \le f(\bar{\mathbf{x}}). \tag{3.22}$$

Let $\boldsymbol{\xi}$ be a subgradient of f at $\bar{\mathbf{x}}$. By the convexity of f, we have

$$f[\bar{\mathbf{x}} + \lambda(\mathbf{x}-\bar{\mathbf{x}})] - f(\bar{\mathbf{x}}) \ge \lambda\boldsymbol{\xi}^t(\mathbf{x}-\bar{\mathbf{x}}).$$

The above inequality, together with (3.20), implies that $\lambda\boldsymbol{\xi}^t(\mathbf{x}-\bar{\mathbf{x}}) \le 0$, and dividing by $\lambda > 0$, the result follows.

Corollary

In addition to the assumptions of the theorem, suppose that f is differentiable. If $\bar{\mathbf{x}} \in S$ is a local optimal solution, $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \le 0$ for all $\mathbf{x} \in S$.

Note that the above result is, in general, necessary but not sufficient for optimality. To illustrate, let $f(x) = x^2$ and $S = \{x : -1 \le x \le 2\}$. The maximum of f over S is equal to 4 and is achieved at $x = 2$. However, at $\bar{x} = 0$, we have $\nabla f(\bar{x}) = 0$ and hence $\nabla f(\bar{x})^t(x-\bar{x}) = 0$ for each $x \in S$. Clearly, the point $\bar{x} = 0$ is not even a local maximum. Referring to Example 3.4.5, discussed earlier, we have two local maxima, (0, 0) and (11/2, 0). Both points satisfy the necessary condition of Theorem 3.4.6. If we are currently at the local optimal point (0, 0), unfortunately no local information exists that will lead us toward the global maximum point (11/2, 0). Also, if we are at the global maximum point (11/2, 0), there is no convenient local criterion that tells us that we are at the optimal point.

Theorem 3.4.7 shows that a convex function achieves a maximum over a compact polyhedral set at an extreme point. This result has been utilized by several computational schemes for solving such problems. We ask the reader to think for a moment about the case when the objective function is linear and, hence, both convex and concave. Theorem 3.4.7 could be extended to the case where the convex feasible region is not polyhedral.

3.4.7 Theorem

Let $f\colon R^n \to R$ be a convex function, and let S be a nonempty compact polyhedral set in R^n. Consider the problem to maximize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. An optimal solution $\bar{\mathbf{x}}$ to the problem then exists, where $\bar{\mathbf{x}}$ is an extreme point of S.

Proof

By Theorem 3.1.3, note that f is continuous. Since S is compact, f assumes a maximum at $\mathbf{x}' \in S$. If $\mathbf{x}'$ is an extreme point of S, the result is at hand. Otherwise, by Theorem 2.6.7, $\mathbf{x}' = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$, where $\sum_{j=1}^{k} \lambda_j = 1$, $\lambda_j > 0$, and $\mathbf{x}_j$ is an extreme point of S for $j = 1,\ldots, k$. By the convexity of f, we have

$$f(\mathbf{x}') = f\left(\sum_{j=1}^{k} \lambda_j \mathbf{x}_j\right) \le \sum_{j=1}^{k} \lambda_j f(\mathbf{x}_j).$$

But since $f(\mathbf{x}') \ge f(\mathbf{x}_j)$ for $j = 1,\ldots, k$, the above inequality implies that $f(\mathbf{x}') = f(\mathbf{x}_j)$ for $j = 1,\ldots, k$. Thus, the extreme points $\mathbf{x}_1,\ldots,\mathbf{x}_k$ are optimal solutions to the problem, and the proof is complete.

3.5 Generalizations of a Convex Functions

In this section we present various types of functions that are similar to convex and concave functions but that share only some of their desirable properties. As we shall learn, many of the results presented later in the book do not require the restrictive assumption of convexity, but rather, the less restrictive assumptions of quasiconvexity, pseudoconvexity, and convexity at a point.

Quasiconvex Functions

Definition 3.5.1 introduces quasiconvex functions. From the definition it is apparent that every convex function is also quasiconvex.

3.5.1 Definition

Let $f: S \to R$, where S is a nonempty convex set in R^n. The function f is said to be *quasiconvex* if for each $\mathbf{x}_1$ and $\mathbf{x}_2 \in S$, the following inequality is true:

$$f[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] \le \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \quad \text{for each } \lambda \in (0, 1).$$

The function f is said to be *quasiconcave* if $-f$ is quasiconvex.

From Definition 3.5.1, a function f is quasiconvex if whenever $f(\mathbf{x}_2) \ge f(\mathbf{x}_1)$, $f(\mathbf{x}_2)$ is greater than or equal to f at all convex combinations of $\mathbf{x}_1$ and $\mathbf{x}_2$. Hence, if f increases from its value at a point along any direction, it must remain nondecreasing in that direction. Therefore, its univariate cross section is either monotone or unimodal (see Exercise 3.57). A function f is quasiconcave if whenever $f(\mathbf{x}_2) \ge f(\mathbf{x}_1)$, f at all convex combinations of $\mathbf{x}_1$ and $\mathbf{x}_2$ is greater than or equal to $f(\mathbf{x}_1)$. Figure 3.10 shows some examples of quasiconvex and quasiconcave functions. We shall concentrate on quasiconvex functions; the reader is advised to draw all the parallel results for quasiconcave functions. A function that is both quasiconvex and quasiconcave is called *quasimonotone* (see Figure 3.10d).

We have learned in Section 3.2 that a convex function can be characterized by the convexity of its epigraph. We now learn that a quasiconvex function can be characterized by the convexity of its level sets. This result is given in Theorem 3.5.2.

3.5.2 Theorem

Let $f: S \to R$ where S is a nonempty convex set in R^n. The function f is quasiconvex if and only if $S_\alpha = \{\mathbf{x} \in S : f(\mathbf{x}) \le \alpha\}$ is convex for each real number α.

(*a*) (*b*) (*c*) (*d*)

Figure 3.10 Quasiconvex and quasiconcave functions: (*a*) quasiconvex, (*b*) quasiconcave, (*c*) neither quasiconvex nor quasiconcave, (*d*) quasimonotone.

Proof

Suppose that f is quasiconvex, and let $\mathbf{x}_1$, $\mathbf{x}_2 \in S_\alpha$. Therefore, $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ and $\max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \le \alpha$. Let $\lambda \in (0,1)$, and let $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. By the convexity of S, $\mathbf{x} \in S$. Furthermore, by the quasiconvexity of f, $f(\mathbf{x}) \le \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \le \alpha$. Hence, $\mathbf{x} \in S_\alpha$ and thus S_α is convex. Conversely, suppose that S_α is convex for each real number α. Let $\mathbf{x}_1$, $\mathbf{x}_2 \in S$. Furthermore, let $\lambda \in (0,1)$ and $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. Note that $\mathbf{x}_1$, $\mathbf{x}_2 \in S_\alpha$ for $\alpha = \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\}$. By assumption, S_α is convex, so that $\mathbf{x} \in S_\alpha$. Therefore, $f(\mathbf{x}) \le \alpha = \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\}$. Hence, f is quasiconvex, and the proof is complete.

The level set S_α defined in Theorem 3.5.2 is sometimes referred to as a *lower-level set,* to differentiate it from the *upper-level set* $\{\mathbf{x} \in S : f(\mathbf{x}) \ge \alpha\}$, which is convex for all $\alpha \in R$ if and only if f is quasiconcave. Also, it can be shown (see Exercise 3.59) that f is quasimonotone if and only if the *level surface* $\{\mathbf{x} \in S : f(\mathbf{x}) = \alpha\}$ is convex for all $\alpha \in R$.

We now give a result analogous to Theorem 3.4.7. Theorem 3.5.3 shows that the maximum of a continuous quasiconvex function over a compact polyhedral set occurs at an extreme point.

3.5.3 Theorem

Let S be a nonempty compact polyhedral set in R^n, and let f: $R^n \to R$ be quasiconvex and continuous on S. Consider the problem to maximize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Then an optimal solution $\bar{\mathbf{x}}$ to the problem exists, where $\bar{\mathbf{x}}$ is an extreme point of S.

Proof

Note that f is continuous on S and hence attains a maximum, say, at $\mathbf{x}' \in S$. If there is an extreme point whose objective is equal to $f(\mathbf{x}')$, the result is at hand. Otherwise, let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be the extreme points of S, and assume that $f(\mathbf{x}') > f(\mathbf{x}_j)$ for $j = 1, \ldots, k$. By Theorem 2.6.7, $\mathbf{x}'$ can be represented as

$$\mathbf{x}' = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$$

$$\sum_{j=1}^{k} \lambda_j = 1$$

$$\lambda_j \ge 0, \qquad j = 1, \ldots, k.$$

Since $f(\mathbf{x}') > f(\mathbf{x}_j)$ for each j, then

$$f(\mathbf{x}') > \max_{1 \le j \le k} f(\mathbf{x}_j) = \alpha. \tag{3.23}$$

Now, consider the set $S_\alpha = \{\mathbf{x} : f(\mathbf{x}) \le \alpha\}$. Note that $\mathbf{x}_j \in S_\alpha$ for j = 1,..., k, and by the quasiconvexity of f, S_α is convex. Hence, $\mathbf{x}' = \sum_{j=1}^k \lambda_j \mathbf{x}_j$ belongs to S_α. This implies that $f(\mathbf{x}') \le \alpha$, which contradicts (3.23). This contradiction shows that $f(\mathbf{x}') = f(\mathbf{x}_j)$ for some extreme point $\mathbf{x}_j$, and the proof is complete.

Differentiable Quasiconvex Functions

The following theorem gives a necessary and sufficient characterization of a differentiable quasiconvex function. (See Appendix B for a second-order characterization in terms of *bordered Hessian determinants*.)

3.5.4 Theorem

Let S be a nonempty open convex set in R^n, and let f: $S \to R$ be differentiable on S. Then f is quasiconvex if and only if either one of the following equivalent statements holds true:

1. If $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ and $f(\mathbf{x}_1) \le f(\mathbf{x}_2)$, $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) \le 0$.
2. If $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ and $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) > 0$, $f(\mathbf{x}_1) > f(\mathbf{x}_2)$.

Proof

Obviously, statements 1 and 2 are equivalent. We shall prove Part 1. Let f be quasiconvex, and let $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ be such that $f(\mathbf{x}_1) \le f(\mathbf{x}_2)$. By the differentiability of f at $\mathbf{x}_2$, for $\lambda \in (0, 1)$, we have

$$f[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] - f(\mathbf{x}_2) = \lambda \nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) + \lambda \|\mathbf{x}_1 - \mathbf{x}_2\| \alpha[\mathbf{x}_2; \lambda(\mathbf{x}_1 - \mathbf{x}_2)],$$

where $\alpha[\mathbf{x}_2; \lambda(\mathbf{x}_1 - \mathbf{x}_2)] \to 0$ as $\lambda \to 0$. By the quasiconvexity of f, we have $f[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] \le f(\mathbf{x}_2)$, and hence the above equation implies that

$$\lambda \nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) + \lambda \|\mathbf{x}_1 - \mathbf{x}_2\| \alpha[\mathbf{x}_2; \lambda(\mathbf{x}_1 - \mathbf{x}_2)] \le 0.$$

Dividing by λ and letting $\lambda \to 0$, we get $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) \le 0$.

Conversely, suppose that $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ and that $f(\mathbf{x}_1) \le f(\mathbf{x}_2)$. We need to show that given Part 1, we have $f[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] \le f(\mathbf{x}_2)$ for each $\lambda \in (0, 1)$. We do this by showing that the set

$$L = \{\mathbf{x} : \mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ \lambda \in (0,1),\ f(\mathbf{x}) > f(\mathbf{x}_2)\}$$

is empty. By contradiction, suppose that there exists an $\mathbf{x}' \in L$. Therefore, $\mathbf{x}' = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\lambda \in (0, 1)$ and $f(\mathbf{x}') > f(\mathbf{x}_2)$. Since f is differentiable, it is continuous, and there must exist a $\delta \in (0, 1)$ such that

$$f[\mu\mathbf{x}' + (1-\mu)\mathbf{x}_2] > f(\mathbf{x}_2) \qquad \text{for each } \mu \in [\delta, 1] \tag{3.24}$$

and $f(\mathbf{x}') > f[\delta\mathbf{x}' + (1-\delta)\mathbf{x}_2]$. By this inequality and the mean value theorem, we must have

$$0 < f(\mathbf{x}') - f[\delta\mathbf{x}' + (1-\delta)\mathbf{x}_2] = (1-\delta)\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}' - \mathbf{x}_2), \tag{3.25}$$

where $\hat{\mathbf{x}} = \hat{\mu}\mathbf{x}' + (1-\hat{\mu})\mathbf{x}_2$ for some $\hat{\mu} \in (\delta, 1)$. From (3.24) it is clear that $f(\hat{\mathbf{x}}) > f(\mathbf{x}_2)$. Dividing (3.25) by $1 - \delta > 0$, it follows that $\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}' - \mathbf{x}_2) > 0$, which in turn implies that

$$\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}_1 - \mathbf{x}_2) > 0. \tag{3.26}$$

But on the other hand, $f(\hat{\mathbf{x}}) > f(\mathbf{x}_2) \ge f(\mathbf{x}_1)$, and $\hat{\mathbf{x}}$ is a convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$, say $\hat{\mathbf{x}} = \hat{\lambda}\mathbf{x}_1 + (1-\hat{\lambda})\mathbf{x}_2$, where $\hat{\lambda} \in (0,1)$. By the assumption of the theorem, $\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}_1 - \hat{\mathbf{x}}) \le 0$, and thus we must have

$$0 \ge \nabla f(\hat{\mathbf{x}})^t(\mathbf{x}_1 - \hat{\mathbf{x}}) = (1-\hat{\lambda})\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}_1 - \mathbf{x}_2).$$

The above inequality is not compatible with (3.26). Therefore, L is empty, and the proof is complete.

To illustrate Theorem 3.5.4, let $f(x) = x^3$. To check its quasiconvexity, suppose that $f(x_1) \le f(x_2)$, that is, $x_1^3 \le x_2^3$. This is true only if $x_1 \le x_2$. Now consider $\nabla f(x_2)(x_1 - x_2) = 3(x_1 - x_2)x_2^2$. Since $x_1 \le x_2$, $3(x_1 - x_2)x_2^2 \le 0$. Therefore, $f(x_1) \le f(x_2)$ implies that $\nabla f(x_2)(x_1 - x_2) \le 0$, and by the theorem we have that f is quasiconvex. As another illustration, let $f(x_1, x_2) = x_1^3 + x_2^3$. Let $\mathbf{x}_1 = (2, -2)^t$ and $\mathbf{x}_2 = (1, 0)^t$. Note that $f(\mathbf{x}_1) = 0$ and $f(\mathbf{x}_2) = 1$, so that $f(\mathbf{x}_1) < f(\mathbf{x}_2)$. But on the other hand, $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) = (3, 0)(1, -2)^t = 3$. By the necessary part of the theorem, f is not quasiconvex. This also shows that the sum of two quasiconvex functions is not necessarily quasiconvex.

Strictly Quasiconvex Functions

Strictly quasiconvex and strictly quasiconcave functions are especially important in nonlinear programming because they ensure that a local minimum and a local maximum over a convex set are, respectively, a global minimum and a global maximum.

3.5.5 Definition

Let $f: S \to R$, where S is a nonempty convex set in R^n. The function f is said to be *strictly quasiconvex* if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$ with $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$, we have

$$f[\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] < \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \qquad \text{for each } \lambda \in (0, 1).$$

The function f is called *strictly quasiconcave* if $-f$ is strictly quasiconvex. Strictly quasiconvex functions are also sometimes referred to as *semi-strictly quasiconvex*, *functionally convex*, or *explicitly quasiconvex*.

Note from Definition 3.5.5 that every convex function is strictly quasiconvex. Figure 3.11 gives examples of strictly quasiconvex and strictly quasiconcave functions. Also, the definition precludes any "flat spots" from occurring anywhere except at extremizing points. This is formalized by the following theorem, which shows that a local minimum of a strictly quasiconvex function over a convex set is also a global minimum. This property is not enjoyed by quasiconvex functions, as seen in Figure 3.10*a*.

3.5.6 Theorem

Let $f: R^n \to R$ be strictly quasiconvex. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where S is a nonempty convex set in R^n. If $\bar{\mathbf{x}}$ is a local optimal solution, $\bar{\mathbf{x}}$ is also a global optimal solution.

Proof

Assume, on the contrary, that there exists an $\hat{\mathbf{x}} \in S$ with $f(\hat{\mathbf{x}}) < f(\bar{\mathbf{x}})$. By the convexity of S, $\lambda\hat{\mathbf{x}} + (1-\lambda)\bar{\mathbf{x}} \in S$ for each $\lambda \in (0, 1)$. Since $\bar{\mathbf{x}}$ is a local minimum by assumption, $f(\bar{\mathbf{x}}) \le f[\lambda\hat{\mathbf{x}} + (1-\lambda)\bar{\mathbf{x}}]$ for all $\lambda \in (0, \delta)$ and for some $\delta \in$

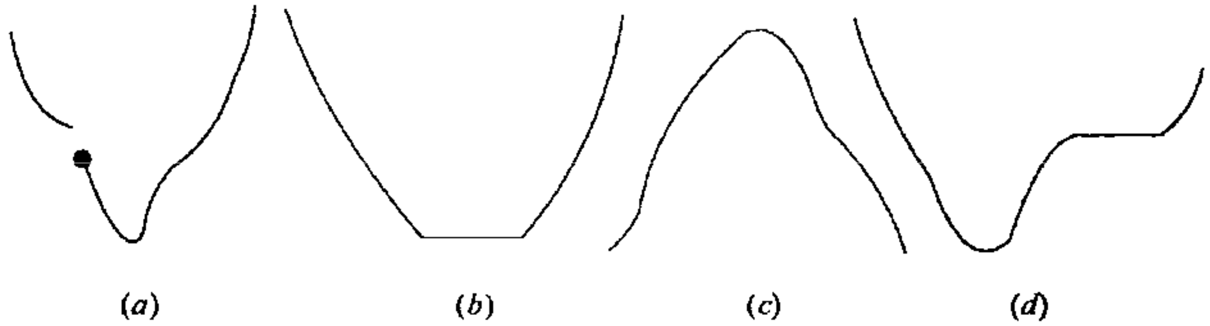


Figure 3.11 Strictly quasiconvex and strictly quasiconcave functions: (*a*) strictly quasiconvex, (*b*) strictly quasiconvex, (*c*) strictly quasiconcave, (*d*) neither strictly quasiconvex nor quasiconcave.

(0, 1). But because f is strictly quasiconvex and $f(\hat{\mathbf{x}}) < f(\overline{\mathbf{x}})$, we have that $f[\lambda\hat{\mathbf{x}} + (1-\lambda)\overline{\mathbf{x}}] < f(\overline{\mathbf{x}})$ for each $\lambda \in (0, 1)$. This contradicts the local optimality of $\overline{\mathbf{x}}$, and the proof is complete.

As seen from Definition 3.1.1, every strictly convex function is indeed a convex function. But every strictly quasiconvex function is not quasiconvex. To illustrate, consider the following function given by Karamardian [1967]:

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0. \end{cases}$$

By Definition 3.5.5, f is strictly quasiconvex. However, f is not quasiconvex, since for $x_1 = 1$ and $x_2 = -1$, $f(x_1) = f(x_2) = 0$, but $f[(1/2)x_1 + (1/2)x_2] = f(0) = 1 > f(x_2)$. If f is lower semicontinuous, however, then as shown below, strict quasiconvexity implies quasiconvexity, as one would usually expect from the word *strict*. (For a definition of lower semicontinuity, refer to Appendix A.)

3.5.7 Lemma

Let S be a nonempty convex set in R^n and let $f: S \to R$ be strictly quasiconvex and lower semicontinuous. Then f is quasiconvex.

Proof

Let $\mathbf{x}_1$ and $\mathbf{x}_2 \in S$. If $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$, then by the strict quasiconvexity of f, we must have $f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] < \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\}$ for each $\lambda \in (0, 1)$. Now, suppose that $f(\mathbf{x}_1) = f(\mathbf{x}_2)$. To show that f is quasiconvex, we need to show that $f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] \le f(\mathbf{x}_1)$ for each $\lambda \in (0, 1)$. By contradiction, suppose that $f[\mu\mathbf{x}_1 + (1-\mu)\mathbf{x}_2] > f(\mathbf{x}_1)$ for some $\mu \in (0, 1)$. Denote $\mu\mathbf{x}_1 + (1-\mu)\mathbf{x}_2$ by $\mathbf{x}$. Since f is lower semicontinuous, there exists a $\lambda \in (0, 1)$ such that

$$f(\mathbf{x}) > f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}] > f(\mathbf{x}_1) = f(\mathbf{x}_2). \tag{3.27}$$

Note that $\mathbf{x}$ can be represented as a convex combination of $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}$ and $\mathbf{x}_2$. Hence, by the strict quasiconvexity of f and since $f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}] > f(\mathbf{x}_2)$, we have $f(\mathbf{x}) < f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}]$, contradicting (3.27). This completes the proof.

Strongly Quasiconvex Functions

From Theorem 3.5.6 it followed that a local minimum of a strictly quasiconvex function over a convex set is also a global optimal solution. However, strict quasiconvexity does not assert uniqueness of the global optimal solution. We shall define here another version of quasiconvexity, called *strong quasiconvexity*, which assures uniqueness of the global minimum when it exists.

3.5.8 Definition

Let S be a nonempty convex set in R^n, and let f: $S \to R$. The function f is said to be *strongly quasiconvex* if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$, with $\mathbf{x}_1 \neq \mathbf{x}_2$, we have

$$f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] < \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\}$$

for each $\lambda \in (0, 1)$. The function f is said to be *strongly quasiconcave* if $-f$ is strongly quasiconvex. (We caution the reader that such a function is sometimes referred to in the literature as being *strictly quasiconvex*, whereas a function satisfying Definition 3.5.5 is called *semi-strictly quasiconvex*. This is done because of Karamardian's example given above and Property 3 below.)

From Definition 3.5.8 and from Definitions 3.1.1, 3.5.1, and 3.5.5, the following statements hold true:

1. Every strictly convex function is strongly quasiconvex.
2. Every strongly quasiconvex function is strictly quasiconvex.
3. Every strongly quasiconvex function is quasiconvex even in the absence of any semicontinuity assumption.

Figure 3.11*a* illustrates a case where the function is both strongly quasiconvex and strictly quasiconvex, whereas the function represented in Figure 3.11*b* is strictly quasiconvex but not strongly quasiconvex. The key to strong quasiconvexity is that it enforces strict unimodality (see Exercise 3.58). This leads to the following property.

3.5.9 Theorem

Let f: $R^n \to R$ be strongly quasiconvex. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where S is a nonempty convex set in R^n. If $\overline{\mathbf{x}}$ is a local optimal solution, $\overline{\mathbf{x}}$ is the unique global optimal solution.

Proof

Since $\overline{\mathbf{x}}$ is a local optimal solution, there exists an ε-neighborhood $N_\varepsilon(\overline{\mathbf{x}})$ around $\overline{\mathbf{x}}$ such that $f(\overline{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in S \cap N_\varepsilon(\overline{\mathbf{x}})$. Suppose, by contradiction to the conclusion of the theorem, that there exists a point $\hat{\mathbf{x}} \in S$ such that $\hat{\mathbf{x}} \neq \overline{\mathbf{x}}$ and $f(\hat{\mathbf{x}}) \leq f(\overline{\mathbf{x}})$. By strong quasiconvexity it follows that

$$f[\lambda\hat{\mathbf{x}} + (1-\lambda)\overline{\mathbf{x}}] < \max\{f(\hat{\mathbf{x}}), f(\overline{\mathbf{x}})\} = f(\overline{\mathbf{x}})$$

for all $\lambda \in (0, 1)$. But for λ small enough, $\lambda\hat{\mathbf{x}} + (1-\lambda)\overline{\mathbf{x}} \in S \cap N_\varepsilon(\overline{\mathbf{x}})$, so that the above inequality violates the local optimality of $\overline{\mathbf{x}}$. This completes the proof.

Pseudoconvex Functions

The astute reader might already have observed that differentiable strongly (or strictly) quasiconvex functions do not share the particular property of convex functions, which says that if $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ at some point $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is a global minimum of f. Figure 3.12c illustrates this fact. This motivates the definition of pseudoconvex functions that share this important property with convex functions, and leads to a generalization of various derivative-based optimality conditions.

3.5.10 Definition

Let S be a nonempty open set in R^n, and let $f: S \to R$ be differentiable on S. The function f is said to be *pseudoconvex* if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$ with $\nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1) \geq 0$, we have $f(\mathbf{x}_2) \geq f(\mathbf{x}_1)$; or equivalently, if $f(\mathbf{x}_2) < f(\mathbf{x}_1)$, $\nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1) < 0$. The function f is said to be *pseudoconcave* if $-f$ is pseudoconvex.

The function f is said to be *strictly pseudoconvex* if for each distinct $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ satisfying $\nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1) \geq 0$, we have $f(\mathbf{x}_2) \geq f(\mathbf{x}_1)$; or equivalently, if for each distinct $\mathbf{x}_1, \mathbf{x}_2 \in S$, $f(\mathbf{x}_2) \leq f(\mathbf{x}_1)$ implies that $\nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1) < 0$. The function f is said to be *strictly pseudoconcave* if $-f$ is strictly pseudoconvex.

Figure 3.12a illustrates a pseudoconvex function. From the definition of pseudoconvexity it is clear that if $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ at any point $\bar{\mathbf{x}}$, $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$ for all $\bar{\mathbf{x}}$; so $\bar{\mathbf{x}}$ is a global minimum for f. Hence, the function in Figure 3.12c is neither pseudoconvex nor pseudoconcave. In fact, the definition asserts that if the directional derivative of f at any point $\mathbf{x}_1$ in the direction $(\mathbf{x}_2 - \mathbf{x}_1)$ is nonnegative, the function values are nondecreasing in that direction (see Exercise 3.69). Furthermore, observe that the pseudoconvex functions shown in Figure 3.12 are also strictly quasiconvex, which is true in general, as shown by Theorem 3.5.11. The reader may note that the function in Figure 3.8c is not pseudoconvex, yet it is strictly quasiconvex.

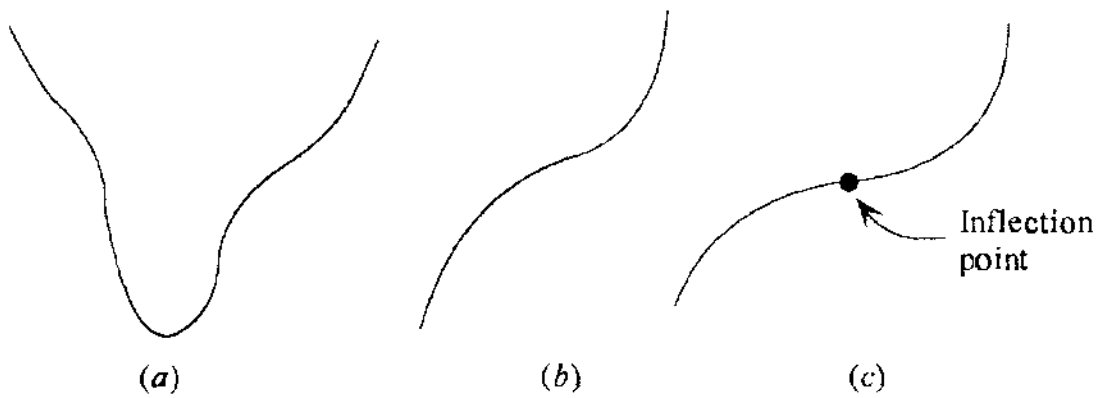


Figure 3.12 Pseudoconvex and pseudoconcave functions: (*a*) pseudoconvex, (*b*) both pseudoconvex and pseudoconcave, (*c*) neither pseudoconvex nor pseudoconcave.

3.5.11 Theorem

Let S be a nonempty open convex set in R^n, and let $f: S \to R$ be a differentiable pseudoconvex function on S. Then f is both strictly quasiconvex and quasiconvex.

Proof

We first show that f is strictly quasiconvex. By contradiction, suppose that there exist $\mathbf{x}_1, \mathbf{x}_2 \in S$ such that $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$ and $f(\mathbf{x}') \geq \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\}$, where $\mathbf{x}' = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\lambda \in (0, 1)$. Without loss of generality, assume that $f(\mathbf{x}_1) < f(\mathbf{x}_2)$, so that

$$f(\mathbf{x}') \geq f(\mathbf{x}_2) > f(\mathbf{x}_1). \tag{3.28}$$

Note, by the pseudoconvexity of f, that $\nabla f(\mathbf{x}')^t(\mathbf{x}_1 - \mathbf{x}') < 0$. Now since $\nabla f(\mathbf{x}')^t$ $(\mathbf{x}_1 - \mathbf{x}') < 0$ and $\mathbf{x}_1 - \mathbf{x}' = -(1-\lambda)(\mathbf{x}_2 - \mathbf{x}')/\lambda$, $\nabla f(\mathbf{x}')^t(\mathbf{x}_2 - \mathbf{x}') > 0$; and hence, by the pseudoconvexity of f, we must have $f(\mathbf{x}_2) \geq f(\mathbf{x}')$. Therefore, by (3.28), we get $f(\mathbf{x}_2) = f(\mathbf{x}')$. Also, since $\nabla f(\mathbf{x}')^t(\mathbf{x}_2 - \mathbf{x}') > 0$, there exists a point $\hat{\mathbf{x}} = \mu\mathbf{x}' + (1-\mu)\mathbf{x}_2$ with $\mu \in (0, 1)$ such that

$$f(\hat{\mathbf{x}}) > f(\mathbf{x}') = f(\mathbf{x}_2).$$

Again, by the pseudoconvexity of f, we have $\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}_2 - \hat{\mathbf{x}}) < 0$. Similarly, $\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}' - \hat{\mathbf{x}}) < 0$. Summarizing, we must have

$$\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}_2 - \hat{\mathbf{x}}) < 0$$

$$\nabla f(\hat{\mathbf{x}})^t(\mathbf{x}' - \hat{\mathbf{x}}) < 0.$$

Note that $\mathbf{x}_2 - \hat{\mathbf{x}} = \mu(\hat{\mathbf{x}} - \mathbf{x}')/(1-\mu)$, and hence the above two inequalities are not compatible. This contradiction shows that f is strictly quasiconvex. By Lemma 3.5.7, then f is also quasiconvex, and the proof is complete.

In Theorem 3.5.12 we see that every strictly pseudoconvex function is strongly quasiconvex.

3.5.12 Theorem

Let S be a nonempty open convex set in R^n, and let $f: S \to R$ be a differentiable strictly pseudoconvex function. Then f is strongly quasiconvex.

Proof

By contradiction, suppose that there exist distinct $\mathbf{x}_1, \mathbf{x}_2 \in S$ and $\lambda \in (0, 1)$ such that $f(\mathbf{x}) \geq \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\}$, where $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. Since $f(\mathbf{x}_1)$

$\le f(\mathbf{x})$, we have, by the strict pseudoconvexity of f, that $\nabla f(\mathbf{x})^t(\mathbf{x}_1 - \mathbf{x}) < 0$ and hence

$$\nabla f(\mathbf{x})^t(\mathbf{x}_1 - \mathbf{x}_2) < 0. \tag{3.29}$$

Similarly, since $f(\mathbf{x}_2) \le f(\mathbf{x})$, we have

$$\nabla f(\mathbf{x})^t(\mathbf{x}_2 - \mathbf{x}_1) < 0. \tag{3.30}$$

The two inequalities (3.29) and (3.30) are not compatible, and hence f is strongly quasiconvex. This completes the proof.

We remark here in connection with Theorems 3.5.11 and 3.5.12, for the special case in which f is quadratic, that f is pseudoconvex if and only if f is strictly quasiconvex, which holds true if and only if f is quasiconvex. Moreover, we also have that f is strictly pseudoconvex if and only if f is strongly quasiconvex. Hence, all these properties become equivalent to each other for quadratic functions (see Exercise 3.55). Also, Appendix B provides a bordered Hessian determinant characterization for checking the pseudoconvexity and the strict pseudoconvexity of quadratic functions.

Thus far we have discussed various types of convexity and concavity. Figure 3.13 summarizes the implications among these types of convexity. These implications either follow from the definitions or from the various results proved in this section. A similar figure can be constructed for the concave case.

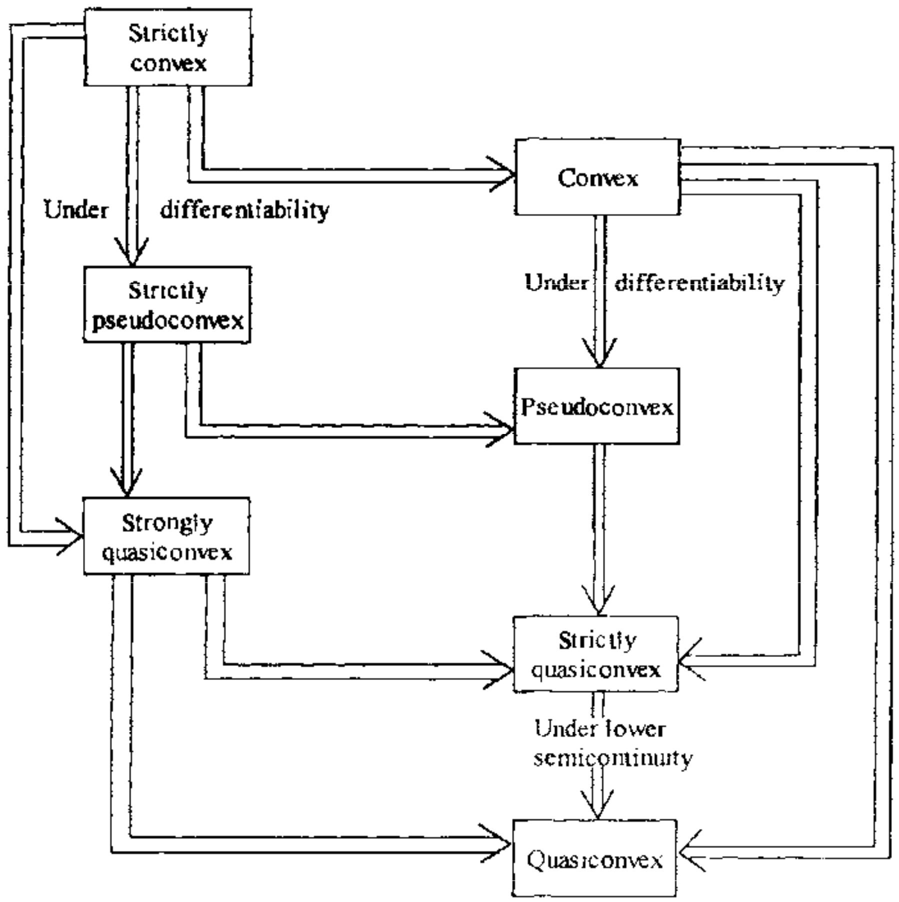


Figure 3.13 Relationship among various types of convexity.

Convexity at a Point

Another useful concept in optimization is the notion of convexity or concavity at a point. In some cases the requirement of a convex or concave function may be too strong and really not essential. Instead, convexity or concavity at a point may be all that is needed.

3.5.13 Definition

Let S be a nonempty convex set in R^n, and let $f: S \to R$. The following are relaxations of various forms of convexity presented in this chapter:

Convexity at $\bar{\mathbf{x}}$. The function f is said to be convex at $\bar{\mathbf{x}} \in S$ if

$$f[\lambda\bar{\mathbf{x}}+(1-\lambda)\mathbf{x}] \le \lambda f(\bar{\mathbf{x}})+(1-\lambda)f(\mathbf{x})$$

for each $\lambda \in (0, 1)$ and each $\mathbf{x} \in S$.

Strict convexity at $\bar{\mathbf{x}}$. The function f is said to be strictly convex at $\bar{\mathbf{x}} \in S$ if

$$f[\lambda\bar{\mathbf{x}}+(1-\lambda)\mathbf{x}] < \lambda f(\bar{\mathbf{x}})+(1-\lambda)f(\mathbf{x})$$

for each $\lambda \in (0, 1)$ and for each $\mathbf{x} \in S$, $\mathbf{x} \ne \bar{\mathbf{x}}$.

Quasiconvexity at $\bar{\mathbf{x}}$. The function f is said to be quasiconvex at $\bar{\mathbf{x}} \in S$ if

$$f[\lambda\bar{\mathbf{x}}+(1-\lambda)\mathbf{x}] \le \max\{f(\mathbf{x}), f(\bar{\mathbf{x}})\}$$

for each $\lambda \in (0, 1)$ and each $\mathbf{x} \in S$.

Strict quasiconvexity at $\bar{\mathbf{x}}$. The function is said to be strictly quasiconvex at $\bar{\mathbf{x}} \in S$ if

$$f[\lambda\bar{\mathbf{x}}+(1-\lambda)\mathbf{x}] < \max\{f(\mathbf{x}), f(\bar{\mathbf{x}})\}$$

for each $\lambda \in (0, 1)$ and each $\mathbf{x} \in S$ such that $f(\mathbf{x}) \ne f(\bar{\mathbf{x}})$.

Strong quasiconvexity at $\bar{\mathbf{x}}$. The function f is said to be strongly quasiconvex at $\bar{\mathbf{x}} \in S$ if

$$f[\lambda\bar{\mathbf{x}}+(1-\lambda)\mathbf{x}] < \max\{f(\mathbf{x}), f(\bar{\mathbf{x}})\}$$

for each $\lambda \in (0,1)$ and each $\mathbf{x} \in S$, $\mathbf{x} \ne \bar{\mathbf{x}}$.

Pseudoconvexity at $\bar{\mathbf{x}}$. The function f is said to be pseudoconvex at $\bar{\mathbf{x}} \in S$ if $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \ge 0$ for $\mathbf{x} \in S$ implies that $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$.

Strict pseudoconvexity at $\bar{\mathbf{x}}$. The function f is said to be strictly pseudoconvex at $\bar{\mathbf{x}} \in S$ if $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \ge 0$ for $\mathbf{x} \in S$, $\mathbf{x} \ne \bar{\mathbf{x}}$, implies that $f(\mathbf{x}) > f(\bar{\mathbf{x}})$.

Various types of concavity at a point can be stated in a similar fashion. Figure 3.14 shows some types of convexity at a point. As the figure suggests, these types of convexity at a point represent a significant relaxation of the concept of convexity.

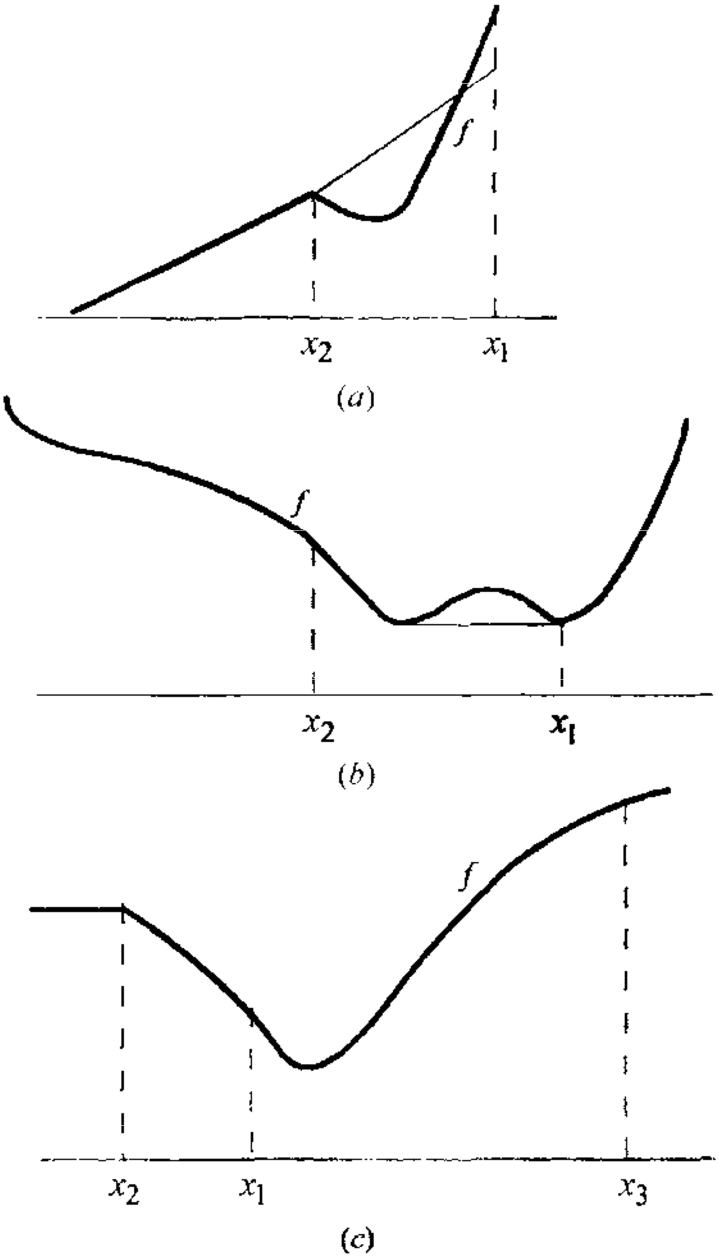


Figure 3.14 Various types of convexity at a point. (*a*) Convexity and strict convexity: f is convex but not strictly convex at x_1; f is both convex and strictly convex at x_2. (*b*) Pseudoconvexity and strict pseudoconvexity: f is pseudoconvex but not strictly pseudoconvex at x_1; f is both pseudoconvex and strictly pseudoconvex at x_2. (*c*) Quasiconvexity, strict quasiconvexity, and strong quasiconvexity: f is quasiconvex but neither strictly quasiconvex nor strongly quasiconvex at x_1; f is both quasiconvex and strictly quasiconvex at x_2 but not strongly quasiconvex at x_2; f is quasiconvex, strictly quasiconvex, and strongly quasiconvex at x_3.

We specify below some important results related to convexity of a function f at a point, where $f: S \to R$ and S is a nonempty convex set in R^n. Of course, not all the results developed throughout this chapter hold true. However, several of these results hold true and are summarized below. The proofs are similar to the corresponding theorems in this chapter.

1. Let f be both convex and differentiable at $\bar{\mathbf{x}}$. Then $f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}})$ for each $\mathbf{x} \in S$. If f is strictly convex, strict inequality holds for $\mathbf{x} \neq \bar{\mathbf{x}}$.
2. Let f be both convex and twice differentiable at $\bar{\mathbf{x}}$. Then the Hessian matrix $\mathbf{H}(\bar{\mathbf{x}})$ is positive semidefinite.
3. Let f be convex at $\bar{\mathbf{x}} \in S$, and let $\bar{\mathbf{x}}$ be an optimal solution to the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Then $\bar{\mathbf{x}}$ is a global optimal solution.
4. Let f be convex and differentiable at $\bar{\mathbf{x}} \in S$. Then $\bar{\mathbf{x}}$ is an optimal solution to the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$ if and only if $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$ for each $\mathbf{x} \in S$. In particular, if $\bar{\mathbf{x}} \in \text{int}\, S$, $\bar{\mathbf{x}}$ is an optimal solution if and only if $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$.
5. Let f be convex and differentiable at $\bar{\mathbf{x}} \in S$. Suppose that $\bar{\mathbf{x}}$ is an optimal solution to the problem to maximize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Then $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$ for each $\mathbf{x} \in S$.
6. Let f be both quasiconvex and differentiable at $\bar{\mathbf{x}}$, and let $\mathbf{x} \in S$ be such that $f(\mathbf{x}) \leq f(\bar{\mathbf{x}})$. Then $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \leq 0$.
7. Suppose that $\bar{\mathbf{x}}$ is a local optimal solution to the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. If f is strictly quasiconvex at $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is a global optimal solution. If f is strongly quasiconvex at $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is the unique global optimal solution.
8. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, and let $\bar{\mathbf{x}} \in S$ be such that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. If f is pseudoconvex at $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is a global optimal solution; and if f is strictly pseudoconvex at $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is the unique global optimal solution.

Exercises

[3.1] Which of the following functions is convex, concave, or neither? Why?

a. $f(x_1, x_2) = 2x_1^2 - 4x_1x_2 - 8x_1 + 3x_2$

b. $f(x_1, x_2) = x_1 e^{-(x_1+3x_2)}$

c. $f(x_1, x_2) = -x_1^2 - 3x_2^2 + 4x_1x_2 + 10x_1 - 10x_2$

d. $f(x_1, x_2, x_3) = 2x_1x_2 + 2x_1^2 + x_2^2 + 2x_3^2 - 5x_1x_3$

e. $f(x_1, x_2, x_3) = -2x_1^2 - 3x_2^2 - 2x_3^2 + 8x_1x_2 + 3x_1x_3 + 4x_2x_3$

[3.2] Over what subset of $\{x : x > 0\}$ is the univariate function $f(x) = e^{-ax^b}$ convex, where $a > 0$ and $b \geq 1$?

[3.3] Prove or disprove concavity of the following function defined over $S = \{(x_1, x_2) : -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$:

$$f(x_1, x_2) = 10 - 3(x_2 - x_1^2)^2.$$

Repeat for a convex set $S \subseteq \{(x_1, x_2) : x_1^2 \geq x_2\}$.

[3.4] Over what domain is the function $f(x) = x^2(x^2 - 1)$ convex? Is it strictly convex over the region(s) specified? Justify your answer.

[3.5] Show that a function f: $R^n \to R$ is affine if and only if f is both convex and concave. [A function f is *affine* if it is of the form $f(\mathbf{x}) = \alpha + \mathbf{c}^t\mathbf{x}$, where a is a scalar and $\mathbf{c}$ is an n-vector.]

[3.6] Let S be a nonempty convex set in R^n, and let f: $S \to R$. Show that f is convex if and only if for any integer $k \geq 2$, the following holds true: $\mathbf{x}_1, \ldots, \mathbf{x}_k \in S$ implies that $f(\sum_{j=1}^k \lambda_j \mathbf{x}_j) \leq \sum_{j=1}^k \lambda_j f(\mathbf{x}_j)$, where $\sum_{j=1}^k \lambda_j = 1$, $\lambda_j \geq 0$ for $j = 1, \ldots, k$.

[3.7] Let S be a nonempty convex set in R^n, and let f: $S \to R$. Show that f is concave if and only if hyp f is convex.

[3.8] Let $f_1, f_2, \ldots, f_k$: $R^n \to R$ be convex functions. Consider the function f defined by $f(\mathbf{x}) = \sum_{j=1}^k \alpha_j f_j(\mathbf{x})$, where $\alpha_j > 0$ for $j = 1, 2, \ldots, k$. Show that f is convex. State and prove a similar result for concave functions.

[3.9] Let $f_1, f_2, \ldots, f_k$: $R^n \to R$ be convex functions. Consider the function f defined by $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})\}$. Show that f is convex. State and prove a similar result for concave functions.

[3.10] Let h: $R^n \to R$ be a convex function, and let g: $R \to R$ be a nondecreasing convex function. Consider the composite function f: $R^n \to R$ defined by $f(\mathbf{x}) = g[h(\mathbf{x})]$. Show that f is convex.

[3.11] Let g: $R^n \to R$ be a concave function, and let f be defined by $f(\mathbf{x}) = 1/g(\mathbf{x})$. Show that f is convex over $S = \{\mathbf{x} : g(\mathbf{x}) > 0\}$. State a symmetric result interchanging the convex and concave functions.

[3.12] Let S be a nonempty convex set in R^n, and let f: $R^n \to R$ be defined as follows:

$$f(\mathbf{y}) = \inf\{\|\mathbf{y} - \mathbf{x}\| : \mathbf{x} \in S\}.$$

Note that $f(\mathbf{y})$ gives the distance from $\mathbf{y}$ to the set S and is called the *distance function*. Prove that f is convex.

[3.13] Let $S = \{(x_1, x_2) : x_1^2 + x_2^2 \le 4\}$. Let f be the distance function defined in Exercise 3.12. Find the function f explicitly.

[3.14] Let S be a nonempty, bounded convex set in R^n, and let f: $R^n \to R$ be defined as follows:

$$f(\mathbf{y}) = \sup\{\mathbf{y}^t\mathbf{x} : \mathbf{x} \in S\}.$$

The function f is called the *support function* of S. Prove that f is convex. Also, show that if $f(\mathbf{y}) = \mathbf{y}^t\overline{\mathbf{x}}$, where $\overline{\mathbf{x}} \in S$, $\overline{\mathbf{x}}$ is a subgradient of f at $\mathbf{y}$.

[3.15] Let $S = A \cup B$, where

$$A = \{(x_1, x_2) : x_1 < 0, x_1^2 + x_2^2 \le 4\}$$

$$B = \{(x_1, x_2) : x_1 \ge 0, -2 \le x_2 \le 2\}.$$

Find the support function defined in Exercise 3.14 explicitly.

[3.16] Let g: $R^m \to R$ be a convex function, and let $\mathbf{h} : R^n \to R^m$ be an affine function of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an $m \times 1$ vector. Then show that the composite function f: $R^n \to R$ defined as $f(\mathbf{x}) = g[\mathbf{h}(\mathbf{x})]$ is a convex function. Also, assuming twice differentiability of g, derive an expression for the Hessian of f.

[3.17] Let F be a *cumulative distribution function* for a random variable b, that is, $F(y) = \text{Prob}(b \le y)$. Show that $\phi(z) = \int_{-\infty}^{z} F(y)\, dy$ is a convex function. Is ϕ convex for any nondecreasing function F?

[3.18] A function f: $R^n \to R$ is called a *gauge function* if it satisfies the following equality:

$$f(\lambda\mathbf{x}) = \lambda f(\mathbf{x}) \qquad \text{for all } \mathbf{x} \in R^n \text{ and all } \lambda \ge 0.$$

Further, a gauge function is said to be *subadditive* if it satisfies the following inequality:

$$f(\mathbf{x}) + f(\mathbf{y}) \ge f(\mathbf{x} + \mathbf{y}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \in R^n.$$

Prove that subadditivity is equivalent to convexity of gauge functions.

[3.19] Let f: $S \to R$ be defined as

$$f(\mathbf{x}) = \frac{(\boldsymbol{\alpha}^t\mathbf{x})^2}{\boldsymbol{\beta}^t\mathbf{x}},$$

where S is a convex subset of R^n, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors in R^n, and where $\boldsymbol{\beta}^t\mathbf{x} > 0$ for all $\mathbf{x} \in S$. Derive an explicit expression for the Hessian of f, and hence verify that f is convex over S.

[3.20] Consider a quadratic function f: $R^n \to R$ and suppose that f is convex on S, where S is a nonempty convex set in R^n. Show that:

a. The function f is convex on $M(S)$, where $M(S)$ is the *affine manifold* containing S defined by $M(S) = \{\mathbf{y} : \mathbf{y} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j,\ \sum_{j=1}^{k} \lambda_j = 1,$ $\mathbf{x}_j \in S$ for all j, for $k \geq 1\}$.
b. The function f is convex on $L(S)$, the *linear subspace* parallel to $M(S)$, defined by $L(S) = \{\mathbf{y} - \mathbf{x} : \mathbf{y} \in M(S)$ and $\mathbf{x} \in S\}$. (This result is credited to Cottle [1967].)

[3.21] Let h: $R^n \to R$ be convex, and let $\mathbf{A}$ be an $m \times n$ matrix. Consider the function h: $R^m \to R$ defined as follows:

$$h(\mathbf{y}) = \inf\{f(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{y}\}.$$

Show that h is convex.

[3.22] Let S be a nonempty convex set in R^n, and let f: $R^n \to R$ and $\mathbf{g}$: $R^n \to R^m$ be convex. Consider the *perturbation function* $\phi : R^m \to R$ defined below:

$$\phi(\mathbf{y}) = \inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{y}, \mathbf{x} \in S\}.$$

a. Prove that ϕ is convex.
b. Show that if $\mathbf{y}_1 \leq \mathbf{y}_2$, $\phi(\mathbf{y}_1) \geq \phi(\mathbf{y}_2)$.

[3.23] Let f: $R^n \to R$ be lower semicontinuous. Show that the level set $S_\alpha = \{\mathbf{x} : f(\mathbf{x}) \leq \alpha\}$ is closed for all $\alpha \in R$.

[3.24] Let f be a convex function on R^n. Prove that the set of subgradients of f at a given point forms a closed convex set.

[3.25] Let f: $R^n \to R$ be convex. Show that $\boldsymbol{\xi}$ is a subgradient of f at $\bar{\mathbf{x}}$ if and only if the hyperplane $\{(\mathbf{x}, y) : y = f(\bar{\mathbf{x}}) + \boldsymbol{\xi}^t(\mathbf{x} - \bar{\mathbf{x}})\}$ supports epi f at $[\bar{\mathbf{x}}, f(\bar{\mathbf{x}})]$. State and prove a similar result for concave functions.

[3.26] Let f: $R^n \to R$ be defined by $f(\mathbf{x}) = \|\mathbf{x}\|$. Prove that subgradients of f are characterized as follows: If $\mathbf{x} = \mathbf{0}$, $\boldsymbol{\xi}$ is a subgradient of f at $\mathbf{x}$ if and only if

$\|\xi\| \le 1$. On the other hand, if $\mathbf{x} \ne \mathbf{0}$, ξ is a subgradient of f at $\mathbf{x}$ if and only if $\|\xi\| = 1$ and $\xi^t \mathbf{x} = \|\mathbf{x}\|$. Use this result to show that f is differentiable at each $\mathbf{x} \ne \mathbf{0}$, and characterize the gradient vector.

[3.27] Let f_1, f_2: $R^n \to R$ be differentiable convex functions. Consider the function f defined by $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$. Let $\bar{\mathbf{x}}$ be such that $f(\bar{\mathbf{x}}) = f_1(\bar{\mathbf{x}}) = f_2(\bar{\mathbf{x}})$. Show that ξ is a subgradient of f at $\bar{\mathbf{x}}$ if and only if

$$\xi = \lambda \nabla f_1(\bar{\mathbf{x}}) + (1-\lambda)\nabla f_2(\bar{\mathbf{x}}), \qquad \text{where } \lambda \in [0, 1].$$

Generalize the result to several convex functions and state a similar result for concave functions.

[3.28] Consider the function θ defined by the following optimization problem for any $\mathbf{u} \ge \mathbf{0}$, where X is a compact polyhedral set.

$$\begin{aligned} \theta(\mathbf{u}) = \ & \text{Minimize } \mathbf{c}^t\mathbf{x} + \mathbf{u}^t(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ & \text{subject to } \mathbf{x} \in X. \end{aligned}$$

a. Show that θ is concave.
b. Characterize the subgradients of θ at any given $\mathbf{u}$.

[3.29] In reference to Exercise 3.28, find the function θ explicitly and describe the set of subgradients at each point $\mathbf{u} \ge \mathbf{0}$ if

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ -1 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$X = \{(x_1, x_2) : 0 \le x_1 \le 3/2,\ 0 \le x_2 \le 3/2\}.$$

[3.30] Consider the function θ defined by the following optimization problem:

$$\begin{aligned} \theta(u_1, u_2) = \ & \text{Minimize } x_1(2 - u_1) + x_2(3 - u_2) \\ & \text{subject to } x_1^2 + x_2^2 \le 4. \end{aligned}$$

a. Show that θ is concave.
b. Evaluate θ at the point (2, 3).
c. Find the collection of subgradients of θ at (2, 3).

[3.31] Let f: $S \to R$, where $S \subseteq R^n$ is a nonempty convex set. Then the ***convex envelope*** of f over S, denoted $f_S(\mathbf{x})$, $\mathbf{x} \in S$, is a convex function such that $f_S(\mathbf{x}) \le f(\mathbf{x})$ for all $\mathbf{x} \in S$; and if g is any other convex function for which $g(\mathbf{x}) \le f(\mathbf{x})$ for all $\mathbf{x} \in S$, $f_S(\mathbf{x}) \ge g(\mathbf{x})$ for all $\mathbf{x} \in S$. Hence f_S is the pointwise supremum over all convex underestimators of f over S. Show that $\min\{f(\mathbf{x}) : \mathbf{x} \in S\} = \min\{f_S(\mathbf{x}) : \mathbf{x} \in S\}$, assuming that the minima exist, and that

$$\{\mathbf{x}^* \in S : f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ for all } \mathbf{x} \in S\}$$
$$\subseteq \{\mathbf{x}^* \in S : f_S(\mathbf{x}^*) \leq f_S(\mathbf{x}) \text{ for all } \mathbf{x} \in S\}.$$

[3.32] Let f: $S \to R$ be a concave function, where $S \subseteq R^n$ is a nonempty polytope with vertices $\mathbf{x}_1, ..., \mathbf{x}_E$. Show that the convex envelope (see Exercise 3.31) of f over S is given by

$$f_S(\mathbf{x}) = \min\left\{\sum_{i=1}^{E} \lambda_i f(\mathbf{x}_i) : \sum_{i=1}^{E} \lambda_i \mathbf{x}_i = \mathbf{x}, \sum_{i=1}^{E} \lambda_i = 1, \lambda_i \geq 0 \text{ for } i = 1, ..., E\right\}.$$

Hence, show that if S is a simplex in R^n, f_S is an affine function that attains the same values as f over all the vertices of S. (This result is due to Falk and Hoffman [1976].)

[3.33] Let f: $S \to R$ and $f_S : S \to R$ be as defined in Exercise 3.31. Show that if f is continuous, the epigraph $\{(\mathbf{x}, y) : y \geq f_S(\mathbf{x}),\ \mathbf{x} \in S,\ y \in R\}$ of f_S over S is the closure of the convex hull of the epigraph $\{(\mathbf{x}, y) : y \geq f(\mathbf{x}),\ \mathbf{x} \in S,\ y \in R\}$ of f over S. Give an example to show that the epigraph of the latter set is not necessarily closed.

[3.34] Let $f(x, y) = xy$ be a bivariate bilinear function, and let S be a polytope in R^2 having no edge with a finite, positive slope. Define $\Lambda = \{(\alpha, \beta, \gamma) \in R^3 : \alpha x_k + \beta y_k + \gamma \leq x_k y_k$ for $k = 1, ..., K\}$, where (x_k, y_k), $k = 1, ..., K$, are the vertices of S. Referring to Exercise 3.31, show that if S is two-dimensional, the set of extreme points $(\alpha_e, \beta_e, \gamma_e)$, $e = 1, ..., E$, of Λ is nonempty and that $f_S(x, y) = \max\{\alpha_e x + \beta_e y + \gamma_e, e = 1, ..., E\}$. On the other hand, if S is one-dimensional and given by the convex hull of (x_1, y_1) and (x_2, y_2), show that there exists a solution $(\alpha_1, \beta_1, \gamma_1)$ to the system $\alpha x_k + \beta y_k + \gamma = x_k y_k$ for $k =$ 1, 2, and in this case, $f_S(x, y) = \alpha_1 x + \beta_1 y + \gamma_1$. Specialize this result to verify that if $S = \{(x, y) : a \leq x \leq b,\ c \leq y \leq d\}$, where $a < b$ and $c < d$, then $f_S(x, y) = \max\{dx + by - bd,\ cx + ay - ac\}$. (This result is due to Sherali and Alameddine [1990].)

[3.35] Consider a triangle S having vertices (0, 1), (2, 0), and (1, 2) and let $f(x, y) = xy$ be a bivariate, bilinear function. Show that the convex envelope f_S of f over S (see Exercise 3.31) is given by

$$f_S(x, y) = \begin{cases} -y + \dfrac{3y^2}{2 - x + y} & \text{for } (x, y) \neq (2, 0) \\ 0 & \text{for } (x, y) = (2, 0) \end{cases} \quad \text{for } (x, y) \in S.$$

Can you generalize your approach to finding the convex envelope of f over a triangle having a single edge that has a finite, positive slope? (This result is due to Sherali and Alameddine [1990].)

[3.36] Let f: $R^n \to R$ be a differentiable function. Show that the gradient vector is given by

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^t.$$

[3.37] Let f: $R^n \to R$, be a differentiable function. The *linear approximation* of f at a given point $\bar{\mathbf{x}}$ is given by

$$f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}).$$

If f is twice differentiable at $\bar{\mathbf{x}}$, the *quadratic approximation* of f at $\bar{\mathbf{x}}$ is given by

$$f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^t \mathbf{H}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}).$$

Let $f(x_1, x_2) = e^{2x_1^2 - x_2^2} - 3x_1 + 5x_2$. Give the linear and quadratic approximations of f at (1, 1). Are these approximations convex, concave, or neither? Why?

[3.38] Consider the function f: $R^n \to R$, and suppose that f is infinitely differentiable. Then show that f is strictly convex if and only if for each $\bar{\mathbf{x}}$ and $\mathbf{d}$ in R^n, the first nonzero derivative term of order greater than or equal to 2 in the Taylor series expansion exists, is of even order, and is positive.

[3.39] Consider the function f: $R^3 \to R$, given by $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A}\mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & \theta \end{bmatrix}.$$

What is the Hessian of f? For what values of θ is f strictly convex?

[3.40] Consider the function $f(x) = x^3$, defined over the set $S = \{x \in R : x \geq 0\}$. Show that f is strictly convex over S. Noting that $f''(0) = 0$ and $f'''(0) = 6$, comment on the application of Theorem 3.3.9.

[3.41] Let $\mathbf{H}$ be an $n \times n$ symmetric, positive semidefinite matrix, and suppose that $\mathbf{x}^t\mathbf{H}\mathbf{x} = 0$ for some $\mathbf{x} \in R^n$. Then show that $\mathbf{H}\mathbf{x} = \mathbf{0}$. (*Hint*: Consider the diagonal of the quadratic form $\mathbf{x}^t\mathbf{H}\mathbf{x}$ via the transformation $\mathbf{x} = \mathbf{Q}\mathbf{y}$, where the columns of $\mathbf{Q}$ are the normalized eigenvectors of $\mathbf{H}$.)

[3.42] Let **H** be an $n \times n$ symmetric matrix. Using the eigenvalue characterization of definiteness, verify that **H** is positive definite if and only if it is positive semidefinite and nonsingular.

[3.43] Suppose that **H** is an $n \times n$ symmetric matrix. Show how Theorem 3.3.12 demonstrates that **H** is positive definite if and only if it can be premultiplied by a series of n lower triangular Gauss–Jordan reduction matrices $\mathbf{L}_1,\ldots,\mathbf{L}_n$ to yield an upper triangular matrix **U** with positive diagonal elements. (Letting $\mathbf{L}^{-1} = \mathbf{L}_n \cdots \mathbf{L}_1$, we obtain **H** = **LU**, where **L** is lower triangular. This is known as the *LU-decomposition* of **H**; see Appendix A.2.) Furthermore, show that **H** is positive definite if and only if there exists a lower triangular matrix **L** with positive diagonal elements such that $\mathbf{H} = \mathbf{L}\mathbf{L}^t$. (This is known as the *Cholesky factorization* of **H**; see Appendix A.2.)

[3.44] Suppose that $S \neq \varnothing$ is closed and convex. Let f: $S \to R$ be differentiable on int S. State if the following are true or false, justifying your answer:

a. If f is convex on S, $f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}})$ for all $\mathbf{x} \in S$, $\bar{\mathbf{x}} \in \text{int } S$.

b. If $f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}})$ for all $\mathbf{x} \in S$ and $\bar{\mathbf{x}} \in \text{int } S$, f is convex on S.

[3.45] Consider the following problem:

$$\begin{aligned} \text{Minimize } & (x_1 - 4)^2 + (x_2 - 6)^2 \\ \text{subject to } & x_2 \geq x_1^2 \\ & x_2 \leq 4. \end{aligned}$$

Write a necessary condition for optimality and verify that it is satisfied by the point (2, 4). Is this the optimal point? Why?

[3.46] Use Theorem 3.4.3 to prove that every local minimum of a convex function over a convex set is also a global minimum.

[3.47] Consider the problem to minimize $\{f(\mathbf{x}) : \mathbf{x} \in S\}$ and suppose that there exists an $\varepsilon > 0$ such that $N_\varepsilon(\bar{\mathbf{x}}) \cap S$ is a convex set and that $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in N_\varepsilon(\bar{\mathbf{x}}) \cap S$.

a. Show that if $\mathbf{H}(\bar{\mathbf{x}})$ is positive definite, $\bar{\mathbf{x}}$ is both a strict and a strong local minimum.

b. Show that if $\bar{\mathbf{x}}$ is a strict local minimum and f is pseudoconvex on $N_\varepsilon(\bar{\mathbf{x}}) \cap S$, $\bar{\mathbf{x}}$ is also a strong local minimum.

[3.48] Let f: $R^n \to R$ be a convex function, and suppose that $f(\mathbf{x} + \lambda \mathbf{d}) \geq f(\mathbf{x})$ for all $\lambda \in (0, \delta)$, where $\delta > 0$. Show that $f(\mathbf{x} + \lambda \mathbf{d})$ is a nondecreasing function of λ. In particular, show that $f(\mathbf{x} + \lambda \mathbf{d})$ is a strictly increasing function of λ if f is strictly convex.

[3.49] Consider the following problem:

$$\begin{aligned}\text{Maximize } & \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \ge \mathbf{0},\end{aligned}$$

where $\mathbf{H}$ is a symmetric negative definite matrix, $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{c}$ is an n-vector, and $\mathbf{b}$ is an m-vector. Write the necessary and sufficient condition for optimality of Theorem 3.4.3, and simplify it using the special structure of this problem.

[3.50] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f: $R^n \to R$ is a differentiable convex function and S is a nonempty convex set in R^n. Prove that $\overline{\mathbf{x}}$ is an optimal solution if and only if $\nabla f(\overline{\mathbf{x}})^t(\mathbf{x}-\overline{\mathbf{x}}) \ge 0$ for each $\mathbf{x} \in S$. State and prove a similar result for the maximization of a concave function. (This result was proved in the text as Corollary 2 to Theorem 3.4.3. In this exercise the reader is asked to give a direct proof without resorting to subgradients.)

[3.51] A vector $\mathbf{d}$ is called a *direction of descent* of f at $\overline{\mathbf{x}}$ if there exists a $\delta > 0$ such that $f(\overline{\mathbf{x}}+\lambda\mathbf{d}) < f(\overline{\mathbf{x}})$ for each $\lambda \in (0,\delta)$. Suppose that f is convex. Show that $\mathbf{d}$ is a direction of descent if and only if $f'(\overline{\mathbf{x}};\mathbf{d}) < 0$. Does the result hold true without the convexity of f?

[3.52] Consider the following problem:

$$\begin{aligned}\text{Maximize } & f(\mathbf{x}) \\ \text{subject to } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \ge \mathbf{0},\end{aligned}$$

where $\mathbf{A}$ is an $m \times n$ matrix with rank m and f is a differentiable convex function. Consider the extreme point $(\mathbf{x}_B^t, \mathbf{x}_N^t) = (\overline{\mathbf{b}}^t, \mathbf{0}^t)$, where $\overline{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b} \ge \mathbf{0}$ and $\mathbf{A} = [\mathbf{B}, \mathbf{N}]$. Decompose $\nabla f(\mathbf{x})$ accordingly into $\nabla_B f(\mathbf{x})$ and $\nabla_N f(\mathbf{x})$. Show that the necessary condition of Theorem 3.4.6 holds true if $\nabla_N f(\mathbf{x})^t - \nabla_B f(\mathbf{x})^t \mathbf{B}^{-1}\mathbf{N} \le \mathbf{0}$. If this condition holds, is it necessarily true that $\mathbf{x}$ is a local maximum? Prove or give a counterexample.

If $\nabla_N f(\mathbf{x})^t - \nabla_B f(\mathbf{x})^t \mathbf{B}^{-1}\mathbf{N} \not\le \mathbf{0}$, choose a positive component j and increase its corresponding nonbasic variable x_j until a new extreme point is reached. Show that this process results in a new extreme point having a larger objective value. Does this method guarantee convergence to a global optimal solution? Prove or give a counterexample.

[3.53] Apply the procedure of Exercise 3.52 to the following problem starting with the extreme point (1/2, 3, 0, 0):

$$\begin{array}{llllll} \text{Maximize } (x_1 & -2)^2 & + (x_2 - 5)^2 & & & \\ \text{subject to } -2x_1 & + \ x_2 & + x_3 & & = 2 \\ \qquad 2x_1 & + 3x_2 & & + x_4 & = 10 \\ \qquad x_1, & x_2, & x_3, & x_4 & \geq 0. \end{array}$$

[3.54] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f: $R^n \to R$ is convex and S is a nonempty convex set in R^n. The cone of feasible directions of S at $\mathbf{x} \in S$ is defined by

$$D = \{\mathbf{d} : \text{there exists a } \delta > 0 \text{ such that } \bar{\mathbf{x}} + \lambda \mathbf{d} \in S \text{ for } \lambda \in (0, \delta)\}.$$

Show that $\bar{\mathbf{x}}$ is an optimal solution to the problem if and only if $f'(\bar{\mathbf{x}}; \mathbf{d}) \geq 0$ for each $\mathbf{d} \in D$. Compare this result with the necessary and sufficient condition of Theorem 3.4.3. Specialize the result to the case where $S = R^n$.

[3.55] Let f: $R^n \to R$ be a quadratic function. Show that f is quasiconvex if and only if it is strictly quasiconvex, which holds true if and only if it is pseudoconvex. Furthermore, show that f is strongly quasiconvex if and only if it is strictly pseudoconvex .

[3.56] Let h: $R^n \to R$ be a quasiconvex function, and let g: $R \to R$ be a nondecreasing function. Then show that the composite function f: $R^n \to R$ defined as $f(\mathbf{x}) = g[h(\mathbf{x})]$ is quasiconvex.

[3.57] Let f: $S \subseteq R \to R$ be a univariate function, where S is some interval on the real line. Define f as *unimodal* on S if there exists an $x^* \in S$ at which f attains a minimum and f is nondecreasing on the interval $\{x \in S : x \geq x^*\}$, whereas it is nonincreasing on the interval $\{x \in S : x \leq x^*\}$. Assuming that f attains a minimum on S, show that f is quasiconvex if and only if it is unimodal on S.

[3.58] Let f: $S \to R$ be a continuous function, where S is a convex subset of R^n. Show that f is quasimonotone if and only if the level surface $\{\mathbf{x} \in S : f(\mathbf{x}) = \alpha\}$ is a convex set for all $\alpha \in R$.

[3.59] Let f: $S \to R$ be a differentiable function, where S is an open, convex subset of R^n. Show that f is quasimonotone if and only if for every $\mathbf{x}_1$ and $\mathbf{x}_2$ in S, $f(\mathbf{x}_1) \geq f(\mathbf{x}_2)$ implies that $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) \geq 0$ and $f(\mathbf{x}_1) \leq f(\mathbf{x}_2)$ implies that $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) \leq 0$. Hence, show that f is quasimonotone if and only if $f(\mathbf{x}_1) \geq f(\mathbf{x}_2)$ implies that $\nabla f(\mathbf{x}_\lambda)^t(\mathbf{x}_1 - \mathbf{x}_2) \geq 0$ for all $\mathbf{x}_1$ and $\mathbf{x}_2$ in S and for all $\mathbf{x}_\lambda = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, where $0 \leq \lambda \leq 1$.

[3.60] Let f: $S \to R$, where f is lower semicontinuous and where S is a convex subset of R^n. Define f as being ***strongly unimodal*** on S if for each $\mathbf{x}_1$ and $\mathbf{x}_2$ in

S for which the function $F(\lambda) = f[\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)]$, $0 \le \lambda \le 1$, attains a minimum at a point $\lambda^* > 0$, we have that $F(0) > F(\lambda) > F(\lambda^*)$ for all $0 < \lambda < \lambda^*$. Show that f is strongly quasiconvex on S if and only if it is strongly unimodal on S (see Exercise 8.10).

[3.61] Let g: $S \to R$ and h: $S \to R$, where S is a nonempty convex set in R^n. Consider the function f: $S \to R$ defined by $f(\mathbf{x}) = g(\mathbf{x})/h(\mathbf{x})$. Show that f is quasiconvex if the following two conditions hold true:

a. g is convex on S, and $g(\mathbf{x}) \ge 0$ for each $\mathbf{x} \in S$.
b. h is concave on S, and $h(\mathbf{x}) > 0$ for each $\mathbf{x} \in S$.

(*Hint:* Use Theorem 3.5.2.)

[3.62] Show that the function f defined in Exercise 3.61 is quasiconvex if the following two conditions hold true:

a. g is convex on S, and $g(\mathbf{x}) \le 0$ for each $\mathbf{x} \in S$.
b. h is convex on S, and $h(\mathbf{x}) > 0$ for each $\mathbf{x} \in S$.

[3.63] Let g: $S \to R$ and h: $S \to R$, where S is a nonempty convex set in R^n. Consider the function f: $S \to R$ defined by $f(\mathbf{x}) = g(\mathbf{x})h(\mathbf{x})$. Show that f is quasiconvex if the following two conditions hold true:

a. g is convex, and $g(\mathbf{x}) \le 0$ for each $\mathbf{x} \in S$.
b. h is concave, and $h(\mathbf{x}) > 0$ for each $\mathbf{x} \in S$.

[3.64] In each of Exercises 3.61, 3.62, and 3.63, show that f is pseudoconvex provided that S is open and that g and h are differentiable.

[3.65] Let $\mathbf{c}_1$, $\mathbf{c}_2$ be nonzero vectors in R^n, and α_1, α_2 be scalars. Let $S = \{\mathbf{x} : \mathbf{c}_2^t\mathbf{x} + \alpha_2 > 0\}$. Consider the function $f : S \to R$ defined as follows:

$$f(\mathbf{x}) = \frac{\mathbf{c}_1^t\mathbf{x} + \alpha_1}{\mathbf{c}_2^t\mathbf{x} + \alpha_2}.$$

Show that f is both pseudoconvex and pseudoconcave. (Functions that are both pseudoconvex and pseudoconcave are called *pseudolinear*.)

[3.66] Consider the quadratic function f: $R^n \to R$ defined by $f(\mathbf{x}) = \mathbf{x}^t\mathbf{H}\mathbf{x}$. The function f is said to be *positive subdefinite* if $\mathbf{x}^t\mathbf{H}\mathbf{x} < 0$ implies that $\mathbf{H}\mathbf{x} \ge \mathbf{0}$ or $\mathbf{H}\mathbf{x} \le \mathbf{0}$ for each $\mathbf{x} \in R^n$. Prove that f is quasiconvex on the *nonnegative orthant*, $R_+^n = \{\mathbf{x} \in R^n : \mathbf{x} \ge \mathbf{0}\}$ if and only if it is positive subdefinite. (This result is credited to Martos [1969].)

[3.67] The function f defined in Exercise 3.66 is said to be *strictly positive subdefinite* if $\mathbf{x}^t\mathbf{H}\mathbf{x} < 0$ implies that $\mathbf{H}\mathbf{x} > \mathbf{0}$ or $\mathbf{H}\mathbf{x} < \mathbf{0}$ for each $\mathbf{x} \in R^n$. Prove that f is pseudoconvex on the nonnegative orthant excluding $\mathbf{x} = \mathbf{0}$ if and only if it is strictly positive subdefinite. (This result is credited to Martos [1969].)

[3.68] Let f: $S \to R$ be a continuously differentiable convex function, where S is some open interval in R. Then show that f is (strictly) pseudoconvex if and only if whenever $f'(\bar{x}) = 0$ for any $\bar{x} \in S$, this implies that $\bar{x}$ is a (strict) local minimum of f on S. Generalize this result to the multivariate case.

[3.69] Let f: $S \to R$ be pseudoconvex, and suppose that for some $\mathbf{x}_1$ and $\mathbf{x}_2$ in R^n, we have $\nabla f(\mathbf{x}_1)^t(\mathbf{x}_2 - \mathbf{x}_1) \geq 0$. Show that the function $F(\lambda)$ = $f[\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)]$ is nondecreasing for $\lambda \geq 0$.

[3.70] Let f: $S \to R$ be a twice differentiable univariate function, where S is some open interval in R. Then show that f is (strictly) pseudoconvex if and only if whenever $f'(\bar{x}) = 0$ for any $\bar{x} \in S$, we have that either $f''(\bar{x}) > 0$ or that $f''(\bar{x}) = 0$ and $\bar{x}$ is a (strict) local minimum of f over S. Generalize this result to the multivariate case.

[3.71] Let $\mathbf{f}$: $R^n \to R^m$ and $\mathbf{g}$: $R^n \to R^k$ be differentiable and convex. Let ϕ: $R^{m+k} \to R$ satisfy the following: If $\mathbf{a}_2 \geq \mathbf{a}_1$ and $\mathbf{b}_2 \geq \mathbf{b}_1$, $\phi(\mathbf{a}_2, \mathbf{b}_2) \geq \phi(\mathbf{a}_1, \mathbf{b}_1)$. Consider the function h: $R^n \to R$ defined by $h(\mathbf{x}) = \phi(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}))$. Show the following:

a. If ϕ is convex, h is convex.
b. If ϕ is pseudoconvex, h is pseudoconvex.
c. If ϕ is quasiconvex, h is quasiconvex.

[3.72] Let g_1, g_2: $R^n \to R$, and let $\alpha \in [0,1]$. Consider the function G_α: $R^n \to R$ defined as

$$G_\alpha(\mathbf{x}) = \frac{1}{2}\left[g_1(\mathbf{x}) + g_2(\mathbf{x}) - \sqrt{g_1^2(\mathbf{x}) + g_2^2(\mathbf{x}) - 2\alpha g_1(\mathbf{x})g_2(\mathbf{x})}\right],$$

where $\sqrt{\ }$ denotes the positive square root.

a. Show that $G_\alpha(\mathbf{x}) \geq 0$ if and only if $g_1(\mathbf{x}) \geq 0$ and $g_2(\mathbf{x}) \geq 0$, that is, minimum $\{g_1(\mathbf{x}), g_2(\mathbf{x})\} \geq 0$.
b. If g_1 and g_2 are differentiable, show that G_α is differentiable at $\mathbf{x}$ for each $\alpha \in [0,1)$ provided that $g_1(\mathbf{x})$, $g_2(\mathbf{x}) \neq 0$.
c. Now suppose that g_1 and g_2 are concave. Show that G_α is concave for α in the interval [0, 1]. Does this result hold true for $\alpha \in (-1,0)$?
d. Suppose that g_1 and g_2 are quasiconcave. Show that G_α is quasiconcave for $\alpha = 1$.
e. Let $g_1(\mathbf{x}) = -x_1^2 - x_2^2 + 4$ and $g_2(\mathbf{x}) = 2x_1 + x_2 - 1$. Obtain an explicit expression for G_α, and verify parts a, b, and c.

This exercise describes a general method for combining two constraints of the form $g_1(\mathbf{x}) \geq 0$ and $g_2(\mathbf{x}) \geq 0$ into an equivalent single constraint of the form $G_\alpha(\mathbf{x}) \geq 0$. This procedure could be applied successively to reduce a problem with several constraints into an equivalent single constrained problem. The procedure is due to Rvačev [1963].

[3.73] Let g_1, g_2: $R^n \to R$, and let $\alpha \in [0,1]$. Consider the function G_α: $R^n \to R$ defined by

$$G_\alpha(\mathbf{x}) = \frac{1}{2}\left[g_1(\mathbf{x}) + g_2(\mathbf{x}) + \sqrt{g_1^2(\mathbf{x}) + g_2^2(\mathbf{x}) - 2\alpha g_1(\mathbf{x}) g_2(\mathbf{x})} \right],$$

where $\sqrt{\ }$ denotes the positive square root.

a. Show that $G_\alpha(\mathbf{x}) \geq 0$ if and only if maximum $\{g_1(\mathbf{x}), g_2(\mathbf{x})\} \geq 0$.

b. If g_1 and g_2 are differentiable, show that G_α is differentiable at $\mathbf{x}$ for each $\alpha \in [0,1)$, provided that $g_1(\mathbf{x})$, $g_2(\mathbf{x}) \neq 0$.

c. Now suppose that g_1 and g_2 are convex. Show that G_α is convex for $\alpha \in [0,1]$. Does the result hold true for $\alpha \in (-1,0)$?

d. Suppose that g_1 and g_2 are quasiconvex. Show that G_α is quasiconvex for $\alpha = 1$.

e. In some optimization problems, the restriction that the variable $x = 0$ or 1 arises. Show that this restriction is equivalent to maximum $\{g_1(x), g_2(x)\} \geq 0$, where $g_1(x) = -x^2$ and $g_2(x) = -(x-1)^2$. Find the function G_α explicitly, and verify statements a, b, and c.

This exercise describes a general method for combining the ***either-or constraints*** of the form $g_1(\mathbf{x}) \geq 0$ or $g_2(\mathbf{x}) \geq 0$ into a single constraint of the form $G_\alpha(\mathbf{x}) \geq 0$, and is due to Rvačev [1963].

Notes and References

In this chapter we deal with the important topic of convex and concave functions. The recognition of these functions is generally traced to Jensen [1905, 1906]. For earlier related works on the subject, see Hadamard [1893] and Hölder [1889].

In Section 3.1, several results related to continuity and directional derivatives of a convex function are presented. In particular, we show that a convex function is continuous on the interior of the domain. See, for example, Rockafellar [1970]. Rockafellar also discusses the ***convex extension*** to R^n of a convex function f: $S \subset R^n \to R$, which takes on finite values over a convex subset S of R^n, by letting $f(\mathbf{x}) = \infty$ for $\mathbf{x} \notin S$. Accordingly, a set of arithmetic operations involving ∞ also needs to be defined. In this case, S is referred to as the ***effective domain*** of f. Also, a ***proper convex function*** is then defined as a convex function for which $f(\mathbf{x}) < \infty$ for at least one point $\mathbf{x}$ and for which $f(\mathbf{x}) > -\infty$ for all $\mathbf{x}$.

In Section 3.2 we discuss subgradients of convex functions. Many of the properties of differentiable convex functions are retained by replacing the gradient vector by a subgradient. For this reason, subgradients have been used frequently in the optimization of nondifferentiable functions. See, for example, Bertsekas [1975], Demyanov and Pallaschke [1985], Demyanov and Vasilev [1985], Held and Karp [1970], Held et al. [1974], Kiwiel [1985], Sherali et al. [2000], Shor [1985], and Wolfe [1976]. (See also, Chapter 8.)

In Section 3.3 we give some properties of differentiable convex functions. For further study of these topics as well as other properties of convex functions, refer to Eggleston [1958], Fenchel [1953], Roberts and Varberg [1973], and Rockafellar [1970]. The superdiagonalization algorithm derived from Theorem 3.3.12 provides an efficient polynomial-time algorithm for checking definiteness properties of matrices. This method is intimately related with LU and Cholesky factorization techniques (see Exercise 3.43, and refer to Section A.2, Fletcher [1985], Luenberger [1973a], and Murty [1983] for further details).

Section 3.4 treats the subject of minima and maxima of convex functions over convex sets. Robinson [1987] discusses the distinction between strict and strong local minima. For general functions, the study of minima and maxima is quite complicated. As shown in Section 3.4, however, every local minimum of a convex function over a convex set is also a global minimum, and the maximum of a convex function over a convex set occurs at an extreme point. For an excellent study of optimization of convex functions, see Rockafellar [1970]. The characterization of the optimal solution set for convex programs is due to Mangasarian [1988]. This paper also extends the results given in Section 3.4 to subdifferentiable convex functions.

In Section 3.5 we examine other classes of functions that are related to convex functions; namely, quasiconvex and pseudoconvex functions. The class of quasiconvex functions was first studied by De Finetti [1949]. Arrow and Enthoven [1961] derived necessary and sufficient conditions for quasiconvexity on the nonnegative orthant assuming twice differentiability. Their results were extended by Ferland [1972]. Note that a local minimum of a quasiconvex function over a convex set is not necessarily a global minimum. This result holds true, however, for a strictly quasiconvex function. Ponstein [1967] introduced the concept of strongly quasiconvex functions, which ensures that the global minimum is unique, a property that is not enjoyed by strictly quasiconvex functions. The notion of pseudoconvexity was introduced by Mangasarian [1965]. The significance of the class of pseudoconvex functions stems from the fact that every point with a zero gradient is a global minimum. Matrix theoretic characterizations (see, e.g., Exercises 3.66 and 3.67) of quadratic pseudoconvex and quasiconvex functions have been presented by Cottle and Ferland [1972] and by Martos [1965, 1967b, 1969, 1975]. For further reading on this topic, refer to Avriel et al. [1988], Fenchel [1953], Greenberg and Pierskalla [1971], Karamardian [1967], Mangasarian [1969a], Ponstein [1967], Schaible [1981a,b], and Schaible and Ziemba [1981]. The last four references give excellent surveys on this topic, and the results of Exercises 3.55 to 3.60 and 3.68 to 3.70 are discussed in detail by Avriel et al. [1988] and Schaible [1981a,b]. Karamardian and Schaible [1990] also present various tests for checking generalized properties for differentiable functions. See also Section B.2.

Exercises 3.31 to 3.34 deal with *convex envelopes* of nonconvex functions. This construct plays an important role in global optimization techniques for nonconvex programming problems. For additional information on this subject, we refer the reader to Al-Khayyal and Falk [1983], Falk [1976], Grotzinger [1985], Horst and Tuy [1990], Pardalos and Rosen [1987], Sherali [1997], and Sherali and Alameddine [1990].

This page is intentionally left blank

Part 2

Optimality Conditions and Duality

This page is intentionally left blank

Chapter 4 The Fritz John and Karush–Kuhn–Tucker Optimality Conditions

In Chapter 3 we derived an optimality condition for a problem of the following form: Minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f is a convex function and S is a convex set. The necessary and sufficient condition for $\bar{\mathbf{x}}$ to solve the problem was shown to be

$$\nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}) \geq 0 \qquad \text{for all } \mathbf{x} \in S.$$

In this chapter the nature of the set S will be specified more explicitly in terms of inequality and/or equality constraints. A set of first-order necessary conditions are derived without any convexity assumptions that are sharper than the above in the sense that they explicitly consider the constraint functions and are more easily verifiable, since they deal with a system of equations. Under suitable convexity assumptions, these necessary conditions are also sufficient for optimality. These optimality conditions lead to *classical* or *direct optimization techniques* for solving unconstrained and constrained problems that construct these conditions and then attempt to find a solution to them. In contrast, we discuss several *indirect methods* in Chapters 8 through 11, which iteratively improve the current solution, converging to a point that can be shown to satisfy these optimality conditions. A discussion of second-order necessary and/or sufficient conditions for unconstrained as well as for constrained problems is also provided.

Readers who are unfamiliar with generalized convexity concepts from Section 3.5 may substitute any references to such properties by related convexity assumptions for ease in reading.

Following is an outline of the chapter.

Section 4.1: Unconstrained Problems We consider briefly optimality conditions for unconstrained problems. First- and second-order conditions are discussed.

Section 4.2: Problems Having Inequality Constraints Both the Fritz John (FJ) and the Karush–Kuhn–Tucker (KKT) conditions for problems having inequality constraints are derived. The nature and value of solutions satisfying these conditions are emphasized.

Section 4.3: Problems Having Inequality and Equality Constraints This section extends the results of the preceding section to problems having both inequality and equality constraints.

Section 4.4: Second-Order Necessary and Sufficient Optimality Conditions for Constrained Problems Similar to the unconstrained case discussed in Section 4.1, we develop second-order necessary and sufficient optimality conditions as an extension to the first-order conditions developed in Sections 4.2 and 4.3 for inequality and equality constrained problems. Many results and algorithms in nonlinear programming assume the existence of a local optimal solution that satisfies the second-order sufficiency conditions.

4.1 Unconstrained Problems

An unconstrained problem is a problem of the form to minimize $f(\mathbf{x})$ without any constraints on the vector $\mathbf{x}$. Unconstrained problems seldom arise in practical applications. However, we consider such problems here because optimality conditions for constrained problems become a logical extension of the conditions for unconstrained problems. Furthermore, as shown in Chapter 9, one strategy for solving a constrained problem is to solve a sequence of unconstrained problems.

We recall below the definitions of local and global minima for unconstrained problems as a special case of Definition 3.4.1, where the set S is replaced by R^n.

4.1.1 Definition

Consider the problem of minimizing $f(\mathbf{x})$ over R^n, and let $\bar{\mathbf{x}} \in R^n$. If $f(\bar{\mathbf{x}}) \le f(\mathbf{x})$ for all $\mathbf{x} \in R^n$, $\bar{\mathbf{x}}$ is called a *global minimum*. If there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ around $\bar{\mathbf{x}}$ such that $f(\bar{\mathbf{x}}) \le f(\mathbf{x})$ for each $\mathbf{x} \in N_\varepsilon(\bar{\mathbf{x}})$, $\bar{\mathbf{x}}$ is called a *local minimum*, while if $f(\bar{\mathbf{x}}) < f(\mathbf{x})$ for all $\mathbf{x} \in N_\varepsilon(\bar{\mathbf{x}})$, $\mathbf{x} \ne \bar{\mathbf{x}}$, for some $\varepsilon > 0$, $\bar{\mathbf{x}}$ is called a *strict local minimum*. Clearly, a global minimum is also a local minimum.

Necessary Optimality Conditions

Given a point $\bar{\mathbf{x}}$ in R^n, we wish to determine, if possible, whether or not the point is a local or a global minimum of a function f. For this purpose we need to characterize a minimizing solution. Fortunately, the differentiability assumption of f provides a means for obtaining this characterization. The corollary to Theorem 4.1.2 gives a first-order necessary condition for $\bar{\mathbf{x}}$ to be a local optimum. Theorem 4.1.3 gives a second-order necessary condition using the Hessian matrix.

4.1.2 Theorem

Suppose that f: $R^n \to R$ is differentiable at $\overline{\mathbf{x}}$. If there is a vector $\mathbf{d}$ such that $\nabla f(\overline{\mathbf{x}})^t\mathbf{d} < 0$, there exists a $\delta > 0$ such that $f(\overline{\mathbf{x}} + \lambda\mathbf{d}) < f(\overline{\mathbf{x}})$ for each $\delta \in (0, \delta)$, so that $\mathbf{d}$ is a *descent direction* of f at $\overline{\mathbf{x}}$.

Proof

By the differentiability of f at $\overline{\mathbf{x}}$, we must have

$$f(\overline{\mathbf{x}} + \lambda\mathbf{d}) = f(\overline{\mathbf{x}}) + \lambda\nabla f(\overline{\mathbf{x}})^t\mathbf{d} + \lambda\|\mathbf{d}\|\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}),$$

where $\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}) \to 0$ as $\lambda \to 0$. Rearranging the terms and dividing by λ, $\lambda \neq 0$, we get

$$\frac{f(\overline{\mathbf{x}} + \lambda\mathbf{d}) - f(\overline{\mathbf{x}})}{\lambda} = \nabla f(\overline{\mathbf{x}})^t\mathbf{d} + \|\mathbf{d}\|\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}).$$

Since $\nabla f(\overline{\mathbf{x}})^t\mathbf{d} < 0$ and $\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}) \to 0$ as $\lambda \to 0$, there exists a $\delta > 0$ such that $\nabla f(\overline{\mathbf{x}})^t\mathbf{d} + \|\mathbf{d}\|\alpha(\overline{\mathbf{x}};\lambda\mathbf{d}) < 0$ for all $\lambda \in (0, \delta)$. The result then follows.

Corollary

Suppose that f: $R^n \to R$ is differentiable at $\overline{\mathbf{x}}$. If $\overline{\mathbf{x}}$ is a local minimum, $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$.

Proof

Suppose that $\nabla f(\overline{\mathbf{x}}) \neq \mathbf{0}$. Then, letting $\mathbf{d} = -\nabla f(\overline{\mathbf{x}})$, we get $\nabla f(\overline{\mathbf{x}})^t\mathbf{d} = -\|\nabla f(\overline{\mathbf{x}})\|^2 < 0$; and by Theorem 4.1.2, there is a $\delta > 0$ such that $f(\overline{\mathbf{x}} + \lambda\mathbf{d}) < f(\overline{\mathbf{x}})$ for $\lambda \in (0, \delta)$, contradicting the assumption that $\overline{\mathbf{x}}$ is a local minimum. Hence, $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$.

The condition above uses the gradient vector whose components are the first partials of f. Hence, it is called a *first-order condition*. Necessary conditions can also be stated in terms of the Hessian matrix H, whose elements are the second partials of f, and are then called *second-order conditions*. One such condition is given below.

4.1.3 Theorem

Suppose that f: $R^n \to R$ is twice differentiable at $\overline{\mathbf{x}}$. If $\overline{\mathbf{x}}$ is a local minimum, $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and $\mathbf{H}(\overline{\mathbf{x}})$ is positive semidefinite.

Proof

Consider an arbitrary direction $\mathbf{d}$. Then from the differentiability of f at $\bar{\mathbf{x}}$, we have

$$f(\bar{\mathbf{x}}+\lambda\mathbf{d}) = f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{d} + \frac{1}{2}\lambda^2\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + \lambda^2\|\mathbf{d}\|^2\alpha(\bar{\mathbf{x}};\lambda\mathbf{d}), \quad (4.1)$$

where $\alpha(\bar{\mathbf{x}};\lambda\mathbf{d}) \to 0$ as $\lambda \to 0$. Since $\bar{\mathbf{x}}$ is a local minimum, from the corollary to Theorem 4.1.2, we have $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. Rearranging the terms in (4.1) and dividing by $\lambda^2 > 0$, we get

$$\frac{f(\bar{\mathbf{x}}+\lambda\mathbf{d}) - f(\bar{\mathbf{x}})}{\lambda^2} = \frac{1}{2}\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + \|\mathbf{d}\|^2\alpha(\bar{\mathbf{x}};\lambda\mathbf{d}). \quad (4.2)$$

Since $\bar{\mathbf{x}}$ is a local minimum, $f(\bar{\mathbf{x}}+\lambda\mathbf{d}) \ge f(\bar{\mathbf{x}})$ for λ sufficiently small. From (4.2) it is thus clear that $(1/2)\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + \|\mathbf{d}\|^2\alpha(\bar{\mathbf{x}};\lambda\mathbf{d}) \ge 0$ for λ sufficiently small. By taking the limit as $\lambda \to 0$, it follows that $\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} \ge 0$; and hence, since $\mathbf{d}$ was arbitrary, $\mathbf{H}(\bar{\mathbf{x}})$ is positive semidefinite.

Sufficient Optimality Conditions

The conditions discussed thus far are necessary conditions; that is, they must be true for every local optimal solution. On the other hand, a point satisfying these conditions need not be a local minimum. Theorem 4.1.4 gives a sufficient condition for a local minimum.

4.1.4 Theorem

Suppose that f: $R^n \to R$ is twice differentiable at $\bar{\mathbf{x}}$. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and $\mathbf{H}(\bar{\mathbf{x}})$ is positive definite, $\bar{\mathbf{x}}$ is a strict local minimum.

Proof

Since f is twice differentiable at $\bar{\mathbf{x}}$, we must have, for each $\bar{\mathbf{x}} \in R^n$,

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^t\mathbf{H}(\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}}) + \|\mathbf{x}-\bar{\mathbf{x}}\|^2\alpha(\bar{\mathbf{x}};\mathbf{x}-\bar{\mathbf{x}}), \quad (4.3)$$

where $\alpha(\bar{\mathbf{x}};\mathbf{x}-\bar{\mathbf{x}}) \to 0$ as $\mathbf{x} \to \bar{\mathbf{x}}$. Suppose, by contradiction, that $\bar{\mathbf{x}}$ is not a strict local minimum; that is, suppose that there exists a sequence $\{\mathbf{x}_k\}$ converging to $\bar{\mathbf{x}}$ such that $f(\mathbf{x}_k) \le f(\bar{\mathbf{x}})$, $\mathbf{x}_k \ne \bar{\mathbf{x}}$, for each k. Considering this sequence, noting that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and that $f(\mathbf{x}_k) \le f(\bar{\mathbf{x}})$, and denoting $(\mathbf{x}_k - \bar{\mathbf{x}})/\|\mathbf{x}_k - \bar{\mathbf{x}}\|$ by $\mathbf{d}_k$, (4.3) then implies that

$$\frac{1}{2}\mathbf{d}_k^t \mathbf{H}(\overline{\mathbf{x}})\mathbf{d}_k + \alpha(\overline{\mathbf{x}};\mathbf{x}_k - \overline{\mathbf{x}}) \le 0 \qquad \text{for each } k. \tag{4.4}$$

But $\|\mathbf{d}_k\| = 1$ for each k; and hence there exists an index set $\mathscr{K}$ such that $\{\mathbf{d}_k\}_{\mathscr{K}}$ converges to $\mathbf{d}$, where $\|\mathbf{d}\| = 1$. Considering this subsequence and the fact that $\alpha(\overline{\mathbf{x}};\mathbf{x}_k - \overline{\mathbf{x}}) \to 0$ as $k \in \mathscr{K}$ approaches ∞, (4.4) implies that $\mathbf{d}^t\mathbf{H}(\overline{\mathbf{x}})\mathbf{d} \le 0$. This contradicts the assumption that $\mathbf{H}(\overline{\mathbf{x}})$ is positive definite since $\|\mathbf{d}\| = 1$. Therefore, $\overline{\mathbf{x}}$ is indeed a strict local minimum.

Essentially, note that assuming f to be twice continuously differentiable, since $\mathbf{H}(\overline{\mathbf{x}})$ is positive definite, we have that $\mathbf{H}(\mathbf{x})$ is positive definite in an ε-neighborhood of $\overline{\mathbf{x}}$, so f is strictly convex in an ε-neighborhood of $\overline{\mathbf{x}}$. Therefore, as follows from Theorem 3.4.2, $\overline{\mathbf{x}}$ is a strict local minimum, that is, it is the unique global minimum over $N_\varepsilon(\overline{\mathbf{x}})$ for some $\varepsilon > 0$. In fact, noting the second part of Theorem 3.4.2, we can conclude that $\overline{\mathbf{x}}$ is also a strong or isolated local minimum in this case.

In Theorem 4.1.5, we show that the necessary condition $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ is also sufficient for $\overline{\mathbf{x}}$ to be a global minimum if f is pseudoconvex at $\overline{\mathbf{x}}$. In particular, if $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and if $\mathbf{H}(\mathbf{x})$ is positive semidefinite for all $\mathbf{x}$, f is convex, and therefore also pseudoconvex. Consequently, $\overline{\mathbf{x}}$ is a global minimum. This is also evident from Theorem 3.3.3 or from Corollary 2 to Theorem 3.4.3.

4.1.5 Theorem

Let $f\colon R^n \to R$ be pseudoconvex at $\overline{\mathbf{x}}$. Then $\overline{\mathbf{x}}$ is a global minimum if and only if $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$.

Proof

By the corollary to Theorem 4.1.2, if $\overline{\mathbf{x}}$ is a global minimum, $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$. Now suppose that $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$, so that $\nabla f(\overline{\mathbf{x}})^t(\mathbf{x} - \overline{\mathbf{x}}) = 0$ for each $\mathbf{x} \in R^n$. By the pseudoconvexity of f at $\overline{\mathbf{x}}$, it then follows that $f(\mathbf{x}) \ge f(\overline{\mathbf{x}})$ for each $\mathbf{x} \in R^n$, and the proof is complete.

Theorem 4.1.5 provides a necessary *and* sufficient optimality condition in terms of the first-order derivative alone when f is pseudoconvex. In a similar manner, we can derive necessary *and* sufficient conditions for local optimality in terms of higher-order derivatives when f is infinitely differentiable, as an extension to the foregoing results. Toward this end, consider the following result for the ***univariate*** case.

4.1.6 Theorem

Let $f: R \to R$ be an infinitely differentiable univariate function. Then $\bar{x} \in R$ is a local minimum if and only if either $f^{(j)}(\bar{x}) = 0$ for all $j = 1, 2, \ldots$, or else there exists an even $n \geq 2$ such that $f^{(n)}(\bar{x}) > 0$ while $f^{(j)}(\bar{x}) = 0$ for all $1 \leq j < n$, where $f^{(j)}$ denotes the jth-order derivative of f.

Proof

We know that $\bar{x}$ is a local minimum of f if and only if $f(\bar{x}+h) - f(\bar{x}) \geq 0$ for all sufficiently small values of $|h|$. Using the infinite Taylor series representation of $f(\bar{x}+h)$, this holds true if and only if

$$hf^{(1)}(\bar{x}) + \frac{h^2}{2!}f^{(2)}(\bar{x}) + \frac{h^3}{3!}f^{(3)}(\bar{x}) + \frac{h^4}{4!}f^{(4)}(\bar{x}) + \cdots \geq 0$$

for all $|h|$ small enough. Similar to the proof of Theorem 3.3.9, it is readily verified that the foregoing inequality holds true if and only if the condition of the theorem is satisfied, and this completes the proof.

Before proceeding, we remark here that for a local maximum, the condition of Theorem 4.1.6 remains the same, except that we require $f^{(n)}(\bar{x}) < 0$ in lieu of $f^{(n)}(\bar{x}) > 0$. Observe also, noting Theorem 3.3.9, that the above result essentially asserts that for the case under discussion, $\bar{x}$ is a local minimum if and only if f is locally convex about $\bar{x}$. This result can be partially extended, at least in theory, to the case of multivariate functions. Toward this end, suppose that $\bar{\mathbf{x}} \in R^n$ is a local minimum for $f: R^n \to R$. Then this holds true if and only if $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) \geq f(\bar{\mathbf{x}})$ for all $\mathbf{d} \in R^n$ and for all sufficiently small values of $|\lambda|$. Assuming f to be infinitely differentiable, this asserts that for all $\mathbf{d} \in R^n$, $\|\mathbf{d}\| = 1$, we must equivalently have

$$f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}}) = \lambda\nabla f(\bar{\mathbf{x}})^t\mathbf{d} + \frac{\lambda^2}{2!}\mathbf{d}^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d}$$

$$+\frac{\lambda^3}{3!}\sum_i\sum_j\sum_k f_{ijk}(\bar{\mathbf{x}})d_i d_j d_k + \cdots \geq 0$$

for all $-\delta \leq \lambda \leq \delta$, for some $\delta > 0$. Consequently, the first nonzero derivative term, if it exists, must correspond to an even power of λ and must be positive in value.

Note that the foregoing concluding statement is *not* sufficient to claim local optimality of $\bar{\mathbf{x}}$. The difficulty is that it might be the case that this statement holds true, implying that for any $\mathbf{d} \in R^n$, $\|\mathbf{d}\| = 1$, we have $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) \geq f(\bar{\mathbf{x}})$ for all $-\delta_{\mathbf{d}} \leq \lambda \leq \delta_{\mathbf{d}}$ for some $\delta_{\mathbf{d}} > 0$, which depends on $\mathbf{d}$, but then, $\delta_{\mathbf{d}}$ might get vanishingly small as $\mathbf{d}$ varies, so that we cannot assert the existence of a $\delta > 0$ such that $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) \geq f(\bar{\mathbf{x}})$ for all $-\delta \leq \lambda \leq \delta$. In this case, by moving along curves instead of along straight lines, improving values of f might be accessible in the immediate neighborhood of $\bar{\mathbf{x}}$. On the other hand, a valid sufficient condition by Theorem 4.1.5 is that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and that f is convex (or pseudoconvex) over an ε-neighborhood about $\bar{\mathbf{x}}$, for some $\varepsilon > 0$. However, this might not be easy to check, and we might need to assess the situation numerically by examining values of the function at perturbations about the point $\bar{\mathbf{x}}$ (refer also to Exercise 4.19).

To illustrate the above point, consider the following example due to the Italian mathematician Peano. Let $f(x_1, x_2) = (x_2^2 - x_1)(x_2^2 - 2x_1) = 2x_1^2 - 3x_1x_2^2 + x_2^4$. Then we have, at $\bar{\mathbf{x}} = (0,0)^t$,

$$\nabla f(\mathbf{0}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{H}(\mathbf{0}) = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}, f_{122}(\mathbf{0}) = f_{212}(\mathbf{0}) = f_{221}(\mathbf{0}) = -6, f_{2222}(\mathbf{0}) = 24,$$

and all other partial derivatives of f of order 3 or higher are zeros. Hence, we obtain by the Taylor series expansion

$$\begin{aligned} f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}}) &= \frac{\lambda^2}{2}(4d_1^2) + \frac{\lambda^3}{6}(-18d_1d_2^2) + \frac{\lambda^4}{24}(24d_2^4) \\ &= 2\lambda^2\left(d_1 - \frac{3\lambda}{4}d_2^2\right)^2 - \frac{1}{8}\lambda^4 d_2^4. \end{aligned}$$

Note that for any $\mathbf{d} = (d_1, d_2)^t$, $\|\mathbf{d}\| = 1$, if $d_1 \neq 0$, the given necessary condition holds true because the second-order term is positive. On the other hand, if $d_1 = 0$, we must have $d_2 \neq 0$, and the condition holds true again because the first nonzero term is of order 4 and is positive. However, $\bar{\mathbf{x}} = (0,0)^t$ is not a local minimum, as evident from Figure 4.1. We have $f(0, 0) = 0$, while there exist negative values of f in any ε-neighborhood about the point (0, 0). In fact, taking $\mathbf{d} = (\sin\theta, \cos\theta)^t$, we have $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}}) = 2\sin^2\theta\lambda^2 - 3\sin\theta\cos^2\theta\lambda^3 + \cos^4\theta\lambda^4$; and for this to be nonnegative for all $-\delta_\theta \leq \lambda \leq \delta_\theta$, $\delta_\theta > 0$, we observe that as $\theta \to 0^+$, we get $\delta_\theta \to 0^+$ as well (see Exercise 4.11), although at $\theta = 0$ we get $\delta_\theta = \infty$. Hence, we cannot derive a $\delta > 0$ such that $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}}) \geq 0$, for all $\mathbf{d} \in R^n$ and $-\delta \leq \lambda \leq \delta$, so $\bar{\mathbf{x}}$ is not a local minimum.

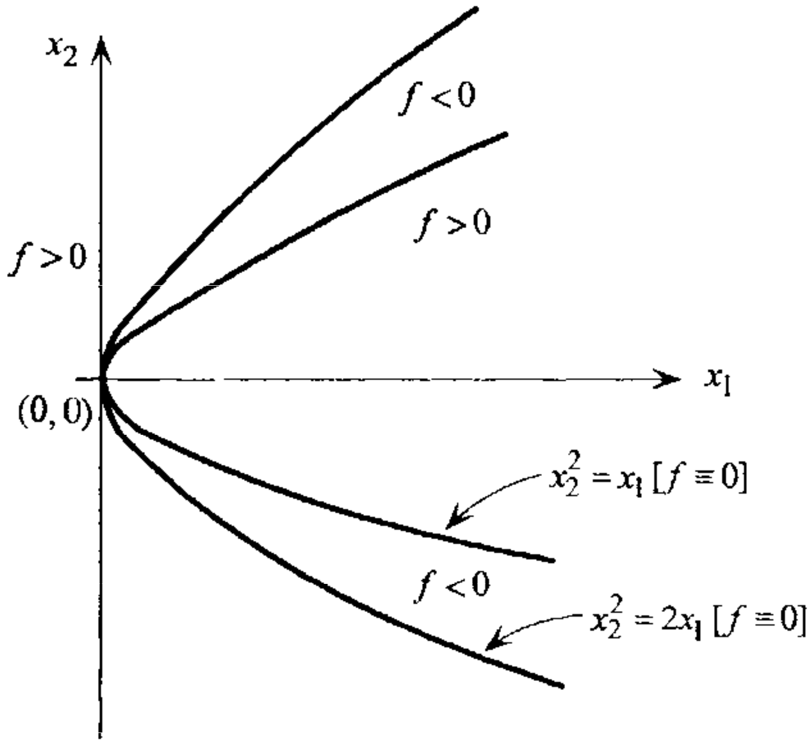


Figure 4.1 Regions of zero, positive, and negative values of $f(x_1, x_2) = (x_2^2 - x_1)\ (x_2^2 - 2x_1)$.

To afford further insight into the multivariate case, let us examine a situation in which f: $R^n \to R$ is twice continuously differentiable, and at a given point $\bar{\mathbf{x}} \in R^n$, we have that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ but $\mathbf{H}(\bar{\mathbf{x}})$ is indefinite. Hence, there exist directions $\mathbf{d}_1$ and $\mathbf{d}_2$ in R^n such that $\mathbf{d}_1^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d}_1 > 0$ and $\mathbf{d}_2^t\mathbf{H}(\bar{\mathbf{x}})\mathbf{d}_2 < 0$. Defining $F_{(\bar{\mathbf{x}};\mathbf{d}_j)}(\lambda) = f(\bar{\mathbf{x}} + \lambda\mathbf{d}_j) \equiv F_{\mathbf{d}_j}(\lambda)$, say, for j = 1, 2, and denoting derivatives by primes, we get

$$F'_{\mathbf{d}_j}(\lambda) = \nabla f(\bar{\mathbf{x}} + \lambda\mathbf{d}_j)^t\mathbf{d}_j \quad \text{and} \quad F''_{\mathbf{d}_j}(\lambda) = \mathbf{d}_j^t\mathbf{H}(\bar{\mathbf{x}} + \lambda\mathbf{d}_j)\mathbf{d}_j \quad \text{for } j = 1, 2.$$

Hence, for j = 1, we have $F'_{\mathbf{d}_1}(0) = 0$, $F''_{\mathbf{d}_1}(0) > 0$; and moreover, by continuity of the second derivative, $F''_{\mathbf{d}_1}(\lambda) > 0$, for $|\lambda|$ sufficiently small. Hence, $F_{\mathbf{d}_1}(\lambda)$ is strictly convex in some ε-neighborhood of $\lambda = 0$, achieving a strict local minimum at $\lambda = 0$. Similarly, for j = 2, noting that $F'_{\mathbf{d}_2}(0) = 0$ and $F''_{\mathbf{d}_2}(0) < 0$, we conclude that $F_{\mathbf{d}_2}(\lambda)$ is strictly concave in some ε-neighborhood of $\lambda = 0$, achieving a strict local maximum at $\lambda = 0$. Hence, as foretold by Theorem 4.1.3, $\bar{\mathbf{x}} = \mathbf{0}$ is neither a local minimum nor a local maximum. Such a point $\bar{\mathbf{x}}$ is called a *saddle point* (or an *inflection point*). Figure 4.2 illustrates the situation. Observe the convex and concave cross sections of the function in the respective directions $\mathbf{d}_1$ and $\mathbf{d}_2$ about the point $\bar{\mathbf{x}}$ at which $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, which gives the function the appearance of a saddle in the vicinity of $\bar{\mathbf{x}}$.

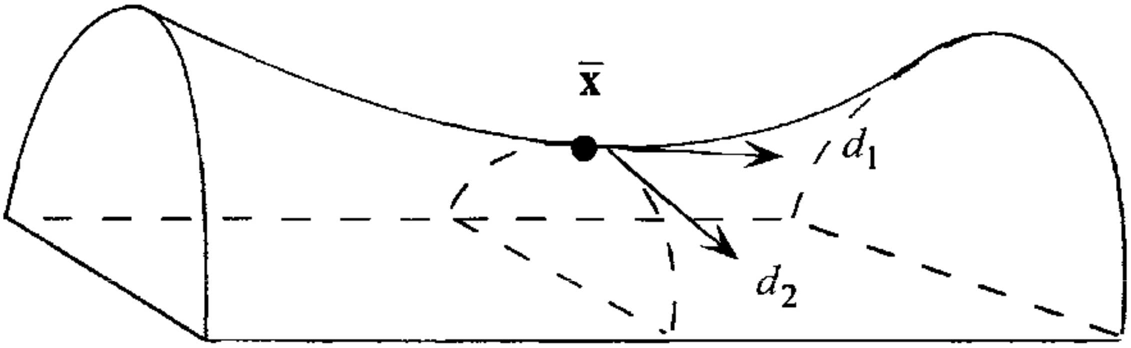


Figure 4.2 Saddle point at $\bar{\mathbf{x}}$.

4.1.7 Examples

Example 1: Univariate Function To illustrate the necessary and sufficient conditions of this section, consider the problem to minimize $f(x) = (x^2 - 1)^3$. First, let us determine the candidate points for optimality satisfying the first-order necessary condition that $\nabla f(x) = 0$. Note that $\nabla f(x) \equiv f'(x) = 6x(x^2 - 1)^2 = 0$ when $x = 0$, 1, or -1. Hence, our candidate points for local optimality are $\bar{x} = 0$, 1, or -1. Now let us examine the second-order derivatives. We have $H(x) = f''(x) = 24x^2(x^2 - 1) + 6(x^2 - 1)^2$, and hence $H(1) = H(-1) = 0$ and $H(0) = 6$. Since H is positive definite at $\bar{x} = 0$, we have by Theorem 4.1.4 that $\bar{x} = 0$ is a strict local minimum. However, at $x = +1$ or -1, H is both positive and negative semidefinite; and although it satisfies the second-order necessary condition of Theorem 4.1.3, this is not sufficient for us to conclude anything about the behavior of f at these points. Hence, we continue and examine the third-order derivative $f'''(x) = 48x(x^2 - 1) + 48x^3 + 24x(x^2 - 1)$. Evaluating this at the two candidate points $\bar{x} = \pm 1$ in question, we obtain $f'''(1) = 48 > 0$ and $f'''(-1) = -48 < 0$. By Theorem 4.1.6 it follows that we have neither a local minimum nor a local maximum at these points, and these points are merely inflection points.

Example 2: Multivariate Function Consider the bivariate function $f(x_1, x_2) = x_1^3 + x_2^3$. Evaluating the gradient and the Hessian of f, we obtain

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 3x_1^2 \\ 3x_2^2 \end{pmatrix} \quad \text{and} \quad \mathbf{H}(\mathbf{x}) = \begin{bmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{bmatrix}.$$

The first-order necessary condition $\nabla f(\bar{\mathbf{x}}) = 0$ yields $\bar{\mathbf{x}} = (0, 0)^t$ as the single candidate point. However, $\mathbf{H}(\bar{\mathbf{x}})$ is the zero matrix; and although it satisfies the second-order necessary condition of Theorem 4.1.3, we need to examine higher-order derivatives to make any conclusive statement about the point $\bar{\mathbf{x}}$. Defining $F_{(\bar{\mathbf{x}};\mathbf{d})}(\lambda) = f(\bar{\mathbf{x}} + \lambda\mathbf{d}) \equiv F_{\mathbf{d}}(\lambda)$, say, we have $F'_{\mathbf{d}}(\lambda) = \nabla f(\bar{\mathbf{x}} + \lambda\mathbf{d})^t\mathbf{d}$, $F''_{\mathbf{d}}(\lambda) =$

$\mathbf{d}^t\mathbf{H}(\overline{\mathbf{x}}+\lambda\mathbf{d})\mathbf{d}$, and $F_{\mathbf{d}}'''(\lambda)=\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} d_i d_j d_k f_{ijk}(\overline{\mathbf{x}}+\lambda\mathbf{d})$. Noting that $f_{111}(\mathbf{x})$ = 6, $f_{222}(\mathbf{x})$ = 6, and $f_{ijk}(\mathbf{x})$ = 0 otherwise, we obtain $F_{\mathbf{d}}'''(0)=6d_1^3+6d_2^3$. Since there exist directions **d** for which the first nonzero derivative term at $\lambda = 0$ is $F_{\mathbf{d}}'''(0)$, which is of odd order, $\overline{\mathbf{x}}=(0,0)^t$ is an inflection point and is therefore neither a local minimum nor a local maximum. In fact, note that $F_{\mathbf{d}}''(\lambda) = 6\lambda(d_1^3 + d_2^3)$ can be made to take on opposite signs about $\lambda = 0$ along any direction **d** for which $d_1^3+d_2^3 \neq 0$; so the function switches from a convex to a concave function, or vice versa, about the point **0** along any direction **d**. Observe also that H is positive semidefinite over $\{\mathbf{x}: x_1 \geq 0,\ x_2 \geq 0\}$; and hence, over this region, the function is convex, yielding $\overline{\mathbf{x}}=(0,0)^t$ as a global minimum. Similarly, $\overline{\mathbf{x}}=(0,0)^t$ is a global maximum over the region $\{\mathbf{x}: x_1 \leq 0,\ x_2 \leq 0\}$.

4.2 Problems Having Inequality Constraints

In this section we first develop a necessary optimality condition for the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$ for a general set S. Later, we let S be more specifically defined as the feasible region of a nonlinear programming problem of the form to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{x} \in X$.

Geometric Optimality Conditions

In Theorem 4.2.2 we develop a necessary optimality condition for the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, using the cone of feasible directions defined below.

4.2.1 Definition

Let S be a nonempty set in R^n, and let $\overline{\mathbf{x}} \in \text{cl}\, S$. The *cone of feasible directions* of S at $\overline{\mathbf{x}}$, denoted by D, is given by

$$D=\{\mathbf{d}:\mathbf{d}\neq\mathbf{0}, \text{ and } \overline{\mathbf{x}}+\lambda\mathbf{d}\in S \text{ for all } \lambda\in(0,\delta) \text{ for some } \delta>0\}.$$

Each nonzero vector $\mathbf{d} \in D$ is called a *feasible direction*. Moreover, given a function f: $R^n \to R$, the *cone of improving directions* at $\overline{\mathbf{x}}$, denoted by F, is given by

$$F=\{\mathbf{d}: f(\overline{\mathbf{x}}+\lambda\mathbf{d})<f(\overline{\mathbf{x}}) \text{ for all } \lambda\in(0,\delta) \text{ for some } \delta>0\}.$$

Each direction $\mathbf{d} \in F$ is called an *improving direction*, or a *descent direction*, of f at $\overline{\mathbf{x}}$.

From the above definitions, it is clear that a small movement from $\overline{\mathbf{x}}$ along a vector $\mathbf{d} \in D$ leads to feasible points, whereas a similar movement along a $\mathbf{d} \in F$ vector leads to solutions of improving objective value. Furthermore, from Theorem 4.1.2, if $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0$, $\mathbf{d}$ is an improving direction; that is, starting from $\bar{\mathbf{x}}$, a small movement along $\mathbf{d}$ will reduce the value of f. As shown in Theorem 4.2.2, if $\bar{\mathbf{x}}$ is a local minimum and if $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0$, $\mathbf{d} \notin D$, that is, a necessary condition for local optimality is that every improving direction is not a feasible direction. This fact is illustrated in Figure 4.3, where the vertices of the cones $F_0 \equiv \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$ and D are translated from the origin to $\bar{\mathbf{x}}$ for convenience.

4.2.2 Theorem

Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, where f: $R^n \to R$ and S is a nonempty set in R^n. Suppose that f is differentiable at a point $\bar{\mathbf{x}} \in S$. If $\bar{\mathbf{x}}$ is a local optimal solution, $F_0 \cap D = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$ and D is the cone of feasible directions of S at $\bar{\mathbf{x}}$. Conversely, suppose that $F_0 \cap D = \varnothing$, f is pseudoconvex at $\bar{\mathbf{x}}$, and that there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$, $\varepsilon > 0$, such that $\mathbf{d} = (\mathbf{x} - \bar{\mathbf{x}}) \in D$ for any $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$. Then $\bar{\mathbf{x}}$ is a local minimum of f.

Proof

By contradiction, suppose that there exists a vector $\mathbf{d} \in F_0 \cap D$. Then, by Theorem 4.1.2, there exists a $\delta_1 > 0$ such that

$$f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}}) \qquad \text{for each } \lambda \in (0, \delta_1). \qquad (4.5a)$$

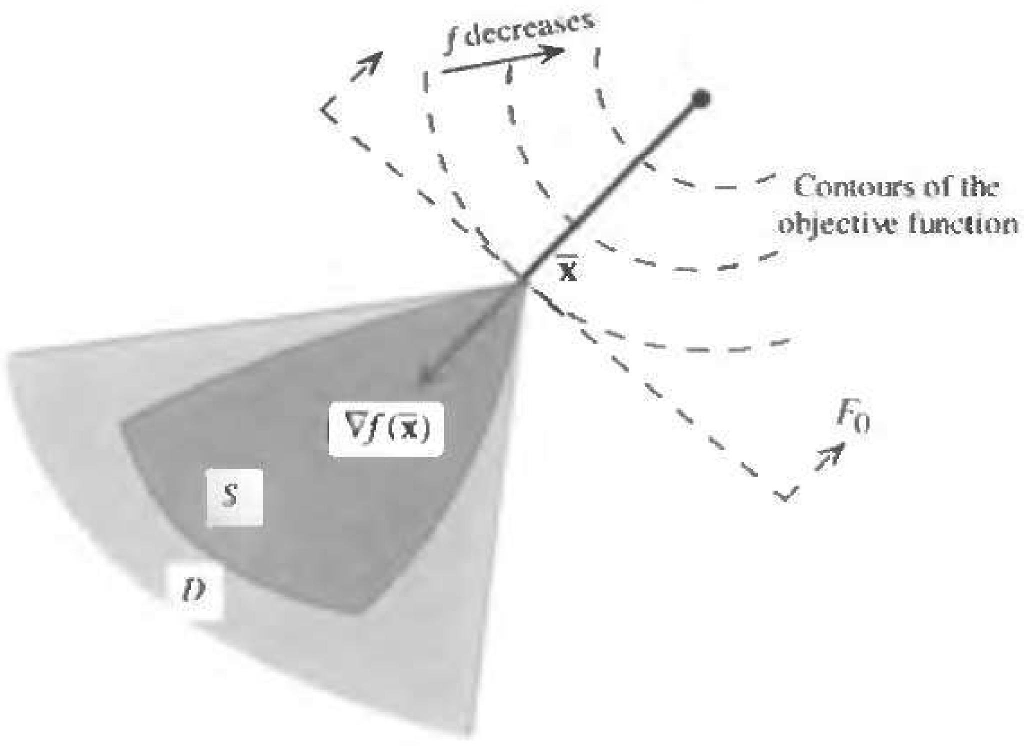


Figure 4.3 Necessary condition $F_0 \cap D = \varnothing$.

Furthermore, by Definition 4.2.1, there exists a $\delta_2 > 0$ such that

$$\bar{\mathbf{x}} + \lambda \mathbf{d} \in S \qquad \text{for each } \lambda \in (0, \delta_2). \tag{4.5b}$$

The assumption that $\bar{\mathbf{x}}$ is a local optimal solution to the problem is not compatible with (4.5). Thus, $F_0 \cap D = \varnothing$.

Conversely, suppose that $F_0 \cap D = \varnothing$ and that the given conditions in the converse statement of the theorem hold true. Then we must have $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$ for all $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$. To understand this, suppose that $f(\hat{\mathbf{x}}) < f(\bar{\mathbf{x}})$ for some $\hat{\mathbf{x}} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$. By the assumption on $S \cap N_\varepsilon(\bar{\mathbf{x}})$, we have $\mathbf{d} = (\hat{\mathbf{x}} - \bar{\mathbf{x}}) \in D$. Moreover, by the pseudoconvexity of f at $\bar{\mathbf{x}}$, we have that $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0$; or else, if $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} \geq 0$, we would obtain $f(\hat{\mathbf{x}}) = f(\bar{\mathbf{x}} + \mathbf{d}) \geq f(\bar{\mathbf{x}})$. We have therefore shown that if $\bar{\mathbf{x}}$ is not a local minimum over $S \cap N_\varepsilon(\bar{\mathbf{x}})$, there exists a direction $\mathbf{d} \in F_0 \cap D$, which is a contradiction. This completes the proof.

Observe that the set F_0 defined in Theorem 4.2.2 provides an algebraic characterization for the set of improving directions F. In fact, we have $F_0 \subseteq F$ in general by Theorem 4.1.2. Also, if $\mathbf{d} \in F$, we must have $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} \leq 0$, or else, analogous to Theorem 4.1.2, $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} > 0$ would imply that d is an ascent direction. Hence, we have

$$F_0 \subseteq F \subseteq F_0' = \{\mathbf{d} \neq \mathbf{0} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} \leq 0\}. \tag{4.6}$$

Note that when $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} = 0$, we are unsure about the behavior of f as we proceed from $\bar{\mathbf{x}}$ along the direction $\mathbf{d}$, unless we know more about the function. For example, it might very well be that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, and there might exist directions of motion that give descent or ascent, or even hold the value of f constant as we move away from $\bar{\mathbf{x}}$. Hence, it is entirely possible to have $F_0 \subset F \subset F_0'$ (see Figure 4.1, for example). However, if f is pseudoconvex, we know that whenever $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} \geq 0$, we have $f(\bar{\mathbf{x}} + \lambda \mathbf{d}) \geq f(\bar{\mathbf{x}})$ for all $\lambda \geq 0$. Hence, if f is pseudoconvex, $\mathbf{d} \in F$ implies that $\mathbf{d} \in F_0$ as well, so from (4.6), we have $F_0 = F$. Similarly, if f is strictly pseudoconcave, we know that whenever $\mathbf{d} \in F_0'$, we have $f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}})$ for all $\lambda > 0$, so we have $\mathbf{d} \in F$ as well. Consequently, we obtain $F = F_0'$ in this case. This establishes the following result, stated in terms of the weaker assumption of pseudoconvexity or strict pseudoconcavity at $\bar{\mathbf{x}}$ itself, rather than everywhere.

4.2.3 Lemma

Consider a differentiable function f: $R^n \to R$, and let F, F_0, F_0' be as defined in Definition 4.2.1, Theorem 4.2.2, and (4.6), respectively. Then we have $F_0 \subseteq F \subseteq F_0'$. Moreover, if f is pseudoconvex at $\bar{\mathbf{x}}$, $F = F_0$, and if f is strictly pseudoconcave at $\bar{\mathbf{x}}$, $F = F_0'$.

We now specify the feasible region S as follows:

$$S = \{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0 \text{ for } i = 1,\ldots,m\},$$

where g_i: $R^n \to R$ for $i = 1,\ldots, m$ and X is a nonempty open set in R^n. This gives us the following inequality constrainted nonlinear programming problem:

$$\begin{aligned} \mathrm{P}: \text{Minimize } & f(\mathbf{x}) \\ \text{subject to } & g_i(\mathbf{x}) \le 0 \qquad \text{for } i = 1,\ldots,m \\ & \mathbf{x} \in X. \end{aligned}$$

Recall that a necessary condition for local optimality at $\bar{\mathbf{x}}$ is that $F_0 \cap D = \varnothing$, where F_0 is an open half-space defined in terms of the gradient vector $\nabla f(\bar{\mathbf{x}})$, and D is the cone of feasible directions, which is not necessarily defined in terms of the gradients of the functions involved. This precludes us from converting the geometric optimality condition $F_0 \cap D = \varnothing$ into a more usable algebraic statement involving equations or inequalities. As Lemma 4.2.4 indicates, we can define an open cone G_0 in terms of the gradients of the binding constraints at $\bar{\mathbf{x}}$ such that $G_0 \subseteq D$. Since $F_0 \cap D = \varnothing$ must hold at $\bar{\mathbf{x}}$ and since $G_0 \subseteq D$, $F_0 \cap G_0 = \varnothing$ is also a necessary optimality condition. Since F_0 and G_0 are both defined in terms of the gradient vectors, we use the condition $F_0 \cap G_0 = \varnothing$ later in the section to develop the optimality conditions credited to Fritz John. With mild additional assumptions, the conditions reduce to the well-known Karush–Kuhn–Tucker (KKT) optimality conditions.

4.2.4 Lemma

Consider the feasible region $S = \{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0 \text{ for } i = 1,\ldots, m\}$, where X is a nonempty open set in R^n and where g_i: $R^n \to R$ for $i = 1,\ldots, m$. Given a feasible point $\bar{\mathbf{x}} \in S$, let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$ be the index set for the *binding*, or *active*, or *tight* constraints, and assume that g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$ are continuous at $\bar{\mathbf{x}}$. Define the sets

$$G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0 \text{ for each } i \in I\}$$

$$G_0' = \{\mathbf{d} \ne \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for each } i \in I\}.$$

Then we have

$$G_0 \subseteq D \subseteq G_0'. \tag{4.7}$$

Moreover, if g_i, $i \in I$, are strictly pseudoconvex at $\overline{\mathbf{x}}$, $D = G_0$; and if g_i, $i \in I$, are pseudoconcave at $\overline{\mathbf{x}}$, $D = G_0'$.

Proof

Let $\mathbf{d} \in G_0$. Since $\overline{\mathbf{x}} \in X$, and X is open, there exists a $\delta_1 > 0$ such that

$$\overline{\mathbf{x}} + \lambda\mathbf{d} \in X \qquad \text{for } \lambda \in (0, \delta_1). \tag{4.8a}$$

Also, since $g_i(\overline{\mathbf{x}}) < 0$ and g_i is continuous at $\overline{\mathbf{x}}$ for $i \notin I$, there exists a $\delta_2 > 0$ such that

$$g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) < 0 \qquad \text{for } \lambda \in (0, \delta_2) \text{ and } i \notin I. \tag{4.8b}$$

Furthermore, since $\mathbf{d} \in G_0$, $\nabla g_i(\overline{\mathbf{x}})^t\mathbf{d} < 0$ for each $i \in I$, and by Theorem 4.1.2, there exists a $\delta_3 > 0$ such that

$$g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) < g_i(\overline{\mathbf{x}}) = 0 \qquad \text{for } \lambda \in (0, \delta_3) \text{ and } i \in I. \tag{4.8c}$$

From (4.8a, b, c), it is clear that points of the form $\overline{\mathbf{x}} + \lambda\mathbf{d}$ are feasible to S for each $\lambda \in (0, \delta)$, where $\delta = \min\{\delta_1, \delta_2, \delta_3\} > 0$. Thus, $\mathbf{d} \in D$, where D is the cone of feasible directions of the feasible region at $\overline{\mathbf{x}}$. We have shown thus far that $\mathbf{d} \in G_0$ implies that $\mathbf{d} \in D$, and hence $G_0 \subseteq D$.

Similarly, if $\mathbf{d} \in D$, we must have $\mathbf{d} \in G_0'$, since otherwise, if $\nabla g_i(\overline{\mathbf{x}})^t\mathbf{d} > 0$ for any $i \in I$, we would obtain via Theorem 4.1.2 that $g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) > g_i(\overline{\mathbf{x}}) = 0$ for all $|\lambda|$ sufficiently small, contradicting the hypotheses that $\mathbf{d} \in D$. Hence, $D \subseteq G_0'$. This establishes (4.7).

Now suppose that g_i, $i \in I$, are strictly pseudoconvex at $\overline{\mathbf{x}}$, and let $\mathbf{d} \in D$. Then we must have $\mathbf{d} \in G_0$ as well, because otherwise, if $\nabla g_i(\overline{\mathbf{x}})^t \mathbf{d} \geq 0$ for any $i \in I$, we would obtain that $g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) > g_i(\overline{\mathbf{x}}) = 0$ for all $\lambda > 0$, contradicting that $\mathbf{d} \in D$. Hence, from (4.7), we get $D = G_0$ in this case.

Finally, suppose that g_i, $i \in I$, are pseudoconcave at $\overline{\mathbf{x}}$, and consider any $\mathbf{d} \in G_0'$. We therefore have $g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) \leq g_i(\overline{\mathbf{x}}) = 0$ for all $\lambda \geq 0$ for each $i \in I$. Moreover, by the continuity of g_i, $i \notin I$, and since X is an open set, we obtain as above that $(\overline{\mathbf{x}} + \lambda\mathbf{d}) \in S$ for all $|\lambda|$ sufficiently small, so $\mathbf{d} \in D$. This establishes that $G_0' \subseteq D$, so from (4.7) we obtain $D = G_0'$ in this case. This completes the proof.

As an illustration, note that in Figure 4.3, $G_0 = D \subset G_0'$, whereas in Figure 2.18 with $\bar{\mathbf{x}} = (0,0)^t$, since the constraint functions are affine, we have $G_0 \subset D = G_0'$.

Lemma 4.2.4 leads directly to the following result.

4.2.5 Theorem

Consider Problem P to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$, where X is a nonempty open set in R^n, f: $R^n \to R$, and g_i: $R^n \to R$, for $i = 1,\ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible point, and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$ are continuous at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is a local optimal solution, $F_0 \cap G_0 = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$ and $G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0 \text{ for each } i \in I\}$. Conversely, if $F_0 \cap G_0 = \varnothing$, and if f is pseudoconvex at $\bar{\mathbf{x}}$ and g_i, $i \in I$, are strictly pseudoconvex over some ε-neighborhood of $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is a local minimum.

Proof

Let $\bar{\mathbf{x}}$ be a local minimum. Then we have the following string of implications via Theorem 4.2.2 and (4.7) of Lemma 4.2.4, which proves the first part of the theorem:

$$\bar{\mathbf{x}} \text{ is a local minimum} \Rightarrow F_0 \cap D = \varnothing \Rightarrow F_0 \cap G_0 = \varnothing. \qquad (4.9a)$$

Conversely, suppose that $F_0 \cap G_0 = \varnothing$ and that f and g_i, $i \in I$, are as specified in the theorem. Then, redefining the feasible region S only in terms of the binding constraints by dropping the nonbinding constraints, we have that $G_0 = D$ by Lemma 4.2.4, so we conclude that $F_0 \cap D = \varnothing$. Furthermore, since the level sets $g_i(\mathbf{x}) \le 0$, for $i \in I$, are convex over some ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ of $\bar{\mathbf{x}}$, $\varepsilon > 0$, it follows that $S \cap N_\varepsilon(\bar{\mathbf{x}})$ is a convex set. Since we also have $F_0 \cap D = \varnothing$ from above, and since f is pseudoconvex at $\bar{\mathbf{x}}$, we conclude from the converse statement of Theorem 4.2.2 that $\bar{\mathbf{x}}$ is a local minimum. This statement continues to hold true by including the nonbinding constraints within S, and this completes the proof.

Observe that under the converse hypothesis of Theorem 4.2.5, and assuming that g_i, $i \notin I$, are continuous at $\bar{\mathbf{x}}$, we have, noting (4.9a),

$$\bar{\mathbf{x}} \text{ is a local minimum} \Leftrightarrow F_0 \cap D = \varnothing \Leftrightarrow F_0 \cap G_0 = \varnothing. \qquad (4.9b)$$

There is a useful insight worth deriving at this point. Note from Definition 4.2.1 that if $\bar{\mathbf{x}}$ is a local minimum, then clearly we must have $F \cap D = \varnothing$. However, the converse is not necessarily true. That is, if $F \cap D = \varnothing$, this does not necessarily imply that $\bar{\mathbf{x}}$ is a local minimum. For example, if

$S = \{\mathbf{x} = (x_1, x_2) : x_2 = x_1^2\}$ and if $f(\mathbf{x}) = x_2$, the point $\bar{\mathbf{x}} = (1,1)^t$ is clearly not a local minimum, since f can be decreased by reducing x_1. However, for the given point $\bar{\mathbf{x}}$, the set $D = \emptyset$, since no *straight-line directions* can lead to feasible solutions, whereas improving feasible solutions are accessible via *curvilinear directions*. Hence, $F \cap D = \emptyset$, but $\bar{\mathbf{x}}$ is not a local minimum. But now, if f is pseudoconvex at $\bar{\mathbf{x}}$, and if there exists an $\varepsilon > 0$ such that for any $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$, we have $\mathbf{d} = (\mathbf{x} - \bar{\mathbf{x}}) \in D$ [as, e.g., if $S \cap N_\varepsilon(\bar{\mathbf{x}})$ is a convex set], $F_0 = F$ by Lemma 4.2.3; and noting (4.9a) and the converse to Theorem 4.2.2, we obtain in this case

$$F \cap D = \emptyset \Leftrightarrow F_0 \cap D = \emptyset \Leftrightarrow \bar{\mathbf{x}} \text{ is a local minimum.}$$

4.2.6 Example

$$\begin{aligned} \text{Minimize } & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{subject to } & x_1^2 + x_2^2 \le 5 \\ & x_1 + x_2 \le 3 \\ & x_1 \ge 0 \\ & x_2 \ge 0. \end{aligned}$$

In this case we let $g_1(\mathbf{x}) = x_1^2 + x_2^2 - 5$, $g_2(\mathbf{x}) = x_1 + x_2 - 3$, $g_3(\mathbf{x}) = -x_1$, $g_4(\mathbf{x}) = -x_2$, and $X = R^2$. Consider the point $\bar{\mathbf{x}} = (9/5, 6/5)^t$, and note that the only binding constraint is $g_2(\mathbf{x}) = x_1 + x_2 - 3$. Also, note that

$$\nabla f(\bar{\mathbf{x}}) = \left(\frac{-12}{5}, \frac{-8}{5}\right)^t \quad \text{and} \quad \nabla g_2(\bar{\mathbf{x}}) = (1,1)^t.$$

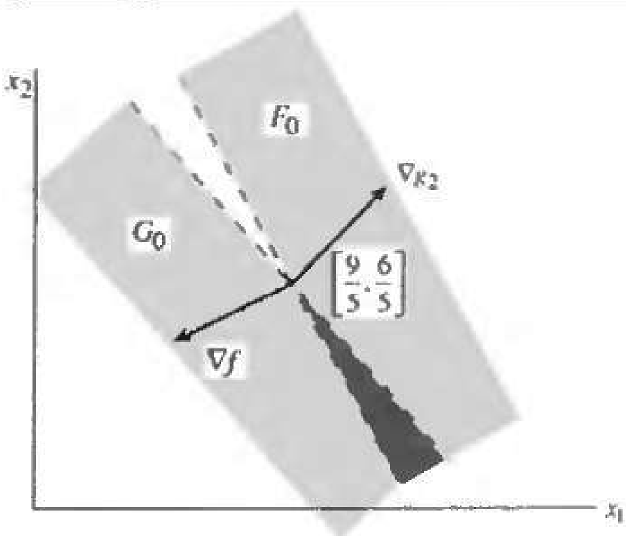


Figure 4.4 $F_0 \cap G_0 \neq \emptyset$ **at a nonoptimal point.**

The sets F_0 and G_0, with the origin translated to $(9/5, 6/5)^t$ for convenience, are shown in Figure 4.4. Since $F_0 \cap G_0 \neq \varnothing$, $\bar{\mathbf{x}} = (9/5, 6/5)^t$ is not a local optimal solution to the above problem.

Now consider the point $\bar{\mathbf{x}} = (2,1)^t$, and note that the first two constraints are binding. The corresponding gradients at this point are

$$\nabla f(\bar{\mathbf{x}}) = (-2,-2)^t, \qquad \nabla g_1(\bar{\mathbf{x}}) = (4,2)^t, \qquad \nabla g_2(\bar{\mathbf{x}}) = (1,1)^t.$$

The sets F_0 and G_0 are shown in Figure 4.5, and indeed, $F_0 \cap G_0 = \varnothing$. Note also that the sufficiency condition of Theorem 4.2.5 is not satisfied because g_2 is not strictly pseudoconvex over any neighborhood about $\bar{\mathbf{x}}$. However, from Figure 4.5 we observe that we also have $F_0 \cap G_0' = \varnothing$ in this case; so by (4.7) we have $F_0 \cap D = \varnothing$. By the converse to Theorem 4.2.2, we can now conclude that $\bar{\mathbf{x}}$ is a local minimum. In fact, since the problem is a convex program with a strictly convex objective function, this in turn implies that $\bar{\mathbf{x}}$ is the unique global minimum.

It might be interesting to note that the utility of Theorem 4.2.5 also depends on how the constraint set is expressed. This is illustrated by Example 4.2.7.

4.2.7 Example

$$\begin{aligned} \text{Minimize} \quad & (x_1-1)^2 + (x_2-1)^2 \\ \text{subject to} \quad & (x_1+x_2-1)^3 \le 0 \\ & x_1 \ge 0 \\ & x_2 \ge 0. \end{aligned}$$

Note that the necessary condition of Theorem 4.2.5 holds true at each feasible point with $x_1 + x_2 = 1$. However, the constraint set can be represented equivalently by

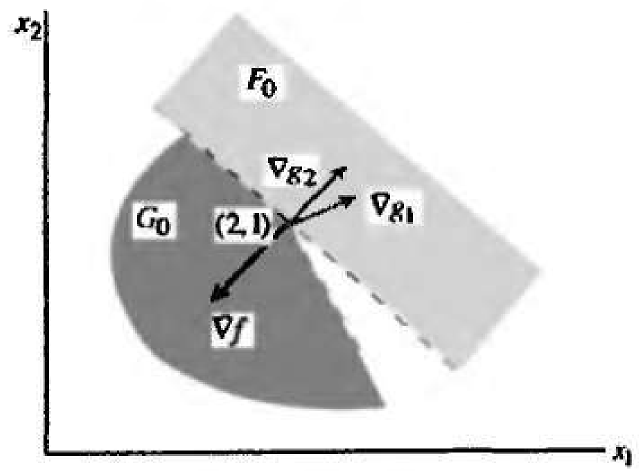


Figure 4.5 $F_0 \cap G_0 = \varnothing$ **at an optimal point.**

$$x_1 + x_2 \le 1$$
$$x_1 \ge 0$$
$$x_2 \ge 0.$$

It can easily be verified that $F_0 \cap G_0 = \varnothing$ is now satisfied only at the point (1/2, 1/2). Moreover, it can also easily be verified that $F_0 \cap G_0' = \varnothing$ in this case, so by (4.7), $F_0 \cap D = \varnothing$. Following the converse to Theorem 4.2.2 and noting the convexity of the feasible region and the strict convexity of the objective function, we can conclude that $\bar{\mathbf{x}} = (1/2, 1/2)^t$ is indeed the unique global minimum to the problem.

There are several cases where the necessary conditions of Theorem 4.2.5 are satisfied trivially by possibly nonoptimal points also. Some of these cases are discussed below.

Suppose that $\bar{\mathbf{x}}$ is a feasible point such that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. Clearly, $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\} = \varnothing$ and hence $F_0 \cap G_0 = \varnothing$. Thus, any point $\bar{\mathbf{x}}$ having $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ satisfies the necessary optimality conditions. Similarly, any point $\bar{\mathbf{x}}$ having $\nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$ for some $i \in I$ will also satisfy the necessary conditions. Now consider the following example with an equality constraint:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g(\mathbf{x}) = 0. \end{aligned}$$

The equality constraint $g(\mathbf{x}) = 0$ could be replaced by the inequality constraints $g_1(\mathbf{x}) \equiv g(\mathbf{x}) \le 0$ and $g_2(\mathbf{x}) \equiv -g(\mathbf{x}) \le 0$. Let $\bar{\mathbf{x}}$ be any feasible point. Then $g_1(\bar{\mathbf{x}}) = g_2(\bar{\mathbf{x}}) = 0$. Note that $\nabla g_1(\bar{\mathbf{x}}) = -\nabla g_2(\bar{\mathbf{x}})$, and therefore there could exist no vector $\mathbf{d}$ such that $\nabla g_1(\bar{\mathbf{x}})^t \mathbf{d} < 0$ and $\nabla g_2(\bar{\mathbf{x}})^t \mathbf{d} < 0$. Therefore, $G_0 = \varnothing$ and hence $F_0 \cap G_0 = \varnothing$. In other words, the necessary condition of Theorem 4.2.5 is satisfied by all feasible solutions and is hence not usable.

Fritz John Optimality Conditions

We now reduce the geometric necessary optimality condition $F_0 \cap G_0 = \varnothing$ to a statement in terms of the gradients of the objective function and of the binding constraints. The resulting optimality conditions, credited to Fritz John [1948], are given below.

4.2.8 Theorem (Fritz John Necessary Conditions)

Let X be a nonempty open set in R^n and let f: $R^n \to R$, and g_i: $R^n \to R$ for $i = 1,\ldots, m$. Consider Problem P to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$

are continuous at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ solves Problem P locally, there exist scalars u_0 and u_i for $i \in I$ such that

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_0, u_i \geq 0 \qquad \text{for } i \in I$$

$$(u_0, \mathbf{u}_I) \neq (0, \mathbf{0}),$$

where $\mathbf{u}_I$ is the vector whose components are u_i for $i \in I$. Furthermore, if g_i for $i \notin I$ are also differentiable at $\bar{\mathbf{x}}$, the foregoing conditions can be written in the following equivalent form:

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1, \ldots, m$$

$$u_0, u_i \geq 0 \qquad \text{for } i = 1, \ldots, m$$

$$(u_0, \mathbf{u}) \neq (0, \mathbf{0}),$$

where $\mathbf{u}$ is the vector whose components are u_i for $i = 1, \ldots, m$.

Proof

Since $\bar{\mathbf{x}}$ solves Problem P locally, by Theorem 4.2.5 there exists no vector $\mathbf{d}$ such that $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0$ and $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0$ for each $i \in I$. Now let $\mathbf{A}$ be the matrix whose rows are $\nabla f(\bar{\mathbf{x}})^t$ and $\nabla g_i(\bar{\mathbf{x}})^t$ for $i \in I$. The necessary optimality condition of Theorem 4.2.5 is then equivalent to the statement that the system $\mathbf{Ad} < \mathbf{0}$ is inconsistent. By Gordan's theorem 2.4.9 there exists a nonzero vector $\mathbf{p} \geq \mathbf{0}$ such that $\mathbf{A}^t \mathbf{p} = \mathbf{0}$. Denoting the components of $\mathbf{p}$ by u_0 and u_i for $i \in I$, the first part of the result follows. The equivalent form of the necessary conditions is readily obtained by letting $u_i = 0$ for $i \notin I$, and the proof is complete.

Pertaining to the conditions of Theorem 4.2.8, the scalars u_0 and u_i for $i = 1, \ldots, m$ are called *Lagrangian*, or *Lagrange, multipliers*. The condition that $\bar{\mathbf{x}}$ be feasible to Problem P is called the *primal feasibility* (PF) *condition*, whereas the requirements $u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$, $(u_0, \mathbf{u}) \geq (0, \mathbf{0})$, and $(u_0, \mathbf{u}) \neq (0, \mathbf{0})$ are sometimes referred to as *dual feasibility* (DF) *conditions*. The condition $u_i g_i(\bar{\mathbf{x}}) = 0$ for $i = 1, \ldots, m$ is called the *complementary slackness* (CS) *condition*. It requires that $u_i = 0$ if the corresponding inequality is nonbinding, that is, if $g_i(\bar{\mathbf{x}}) < 0$. Similarly, it permits $u_i > 0$ only for those constraints that are binding.

Together, the PF, DF, and the CS conditions are called the ***Fritz John* (FJ) *optimality conditions.*** Any point $\bar{\mathbf{x}}$ for which there exist Lagrangian multipliers $(\bar{u}_0, \bar{\mathbf{u}})$ such that $(\bar{\mathbf{x}}, \bar{u}_0, \bar{\mathbf{u}})$ satisfy the FJ conditions is called a ***Fritz John point.*** The Fritz John conditions can also be written in vector notation as follows, in addition to the PF requirement:

$$
\begin{aligned}
u_0\nabla f(\bar{\mathbf{x}})+\nabla \mathbf{g}(\bar{\mathbf{x}})^t\mathbf{u} &= \mathbf{0}\\
\mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}) &= 0\\
(u_0,\mathbf{u}) &\geq (0,\mathbf{0})\\
(u_0,\mathbf{u}) &\neq (0,\mathbf{0}).
\end{aligned}
$$

Here, $\nabla \mathbf{g}(\bar{\mathbf{x}})$ is an $m \times n$ *Jacobian matrix* whose ith row is $\nabla g_i(\bar{\mathbf{x}})^t$, and $\mathbf{u}$ is an m-vector denoting the Lagrangian multipliers.

4.2.9 Example

$$
\begin{aligned}
\text{Minimize } & (x_1-3)^2 + (x_2-2)^2\\
\text{subject to } & x_1^2+x_2^2 \leq 5\\
& x_1+2x_2 \leq 4\\
& -x_1 \leq 0\\
& -x_2 \leq 0.
\end{aligned}
$$

The feasible region for the above problem is illustrated in Figure 4.6. We now verify that the Fritz John conditions hold true at the optimal point (2, 1). First, note that the set of binding constraints I at $\bar{\mathbf{x}} = (2,1)^t$ is given by $I = \{1, 2\}$. Thus, the Lagrangian multipliers u_3 and u_4 associated with $-x_1 \leq 0$ and $-x_2 \leq 0$, respectively, are equal to zero. Note that

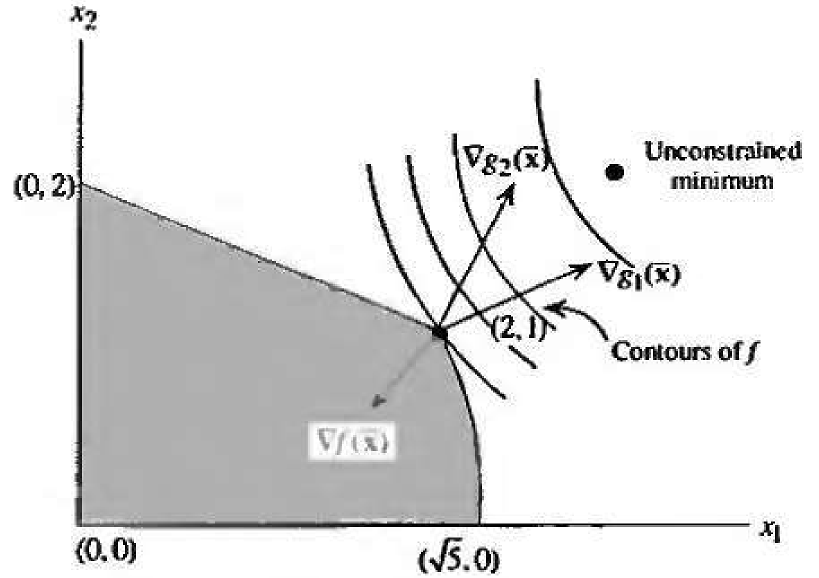


Figure 4.6 Feasible Region in Example 4.2.9.

$$\nabla f(\bar{\mathbf{x}}) = (-2,-2)^t, \qquad \nabla g_1(\bar{\mathbf{x}}) = (4,2)^t, \qquad \nabla g_2(\bar{\mathbf{x}}) = (1,2)^t.$$

Hence, to satisfy the Fritz John conditions, we now need a nonzero vector $(u_0, u_1, u_2) \geq \mathbf{0}$ satisfying

$$u_0\begin{pmatrix}-2\\-2\end{pmatrix} + u_1\begin{pmatrix}4\\2\end{pmatrix} + u_2\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

This implies that $u_1 = u_0/3$ and $u_2 = 2u_0/3$. Taking u_1 and u_2 as such for any $u_0 > 0$, we satisfy the FJ conditions. As another illustration, let us check whether the point $\hat{\mathbf{x}} = (0,0)^t$ is a FJ point. Here the set of binding constraints is $I = \{3, 4\}$, and thus $u_1 = u_2 = 0$. Note that

$$\nabla f(\hat{\mathbf{x}}) = (-6,-4)^t, \qquad \nabla g_3(\hat{\mathbf{x}}) = (-1,0)^t, \qquad \nabla g_4(\hat{\mathbf{x}}) = (0,-1)^t.$$

Also, note that the DF condition

$$u_0\begin{pmatrix}-6\\-4\end{pmatrix} + u_3\begin{pmatrix}-1\\0\end{pmatrix} + u_4\begin{pmatrix}0\\-1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

holds true if and only if $u_3 = -6u_0$ and $u_4 = -4u_0$. If $u_0 > 0$, u_3 and u_4 are negative, contradicting the nonnegativity restrictions. If, on the other hand, $u_0 = 0$, then $u_3 = u_4 = 0$, which contradicts the stipulation that the vector (u_0, u_3, u_4) is nonzero. Thus, the Fritz John conditions do not hold true at $\hat{\mathbf{x}} = (0,0)^t$, which also shows that the origin is not a local optimal point.

4.2.10 Example

Consider the following problem from Kuhn and Tucker [1951]:

$$\begin{aligned}
&\text{Minimize } -x_1 \\
&\text{subject to } x_2 - (1-x_1)^3 \leq 0 \\
&\qquad\qquad\qquad -x_2 \leq 0.
\end{aligned}$$

The feasible region is illustrated in Figure 4.7. We now verify that the Fritz John conditions indeed hold true at the optimal point $\bar{\mathbf{x}} = (1,0)^t$. Note that the set of binding constraints at $\bar{\mathbf{x}}$ is given by $I = \{1, 2\}$. Also,

$$\nabla f(\bar{\mathbf{x}}) = (-1,0)^t, \qquad \nabla g_1(\bar{\mathbf{x}}) = (0,1)^t, \qquad \nabla g_2(\bar{\mathbf{x}}) = (0,-1)^t.$$

The DF condition

$$u_0\begin{pmatrix}-1\\0\end{pmatrix} + u_1\begin{pmatrix}0\\1\end{pmatrix} + u_2\begin{pmatrix}0\\-1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

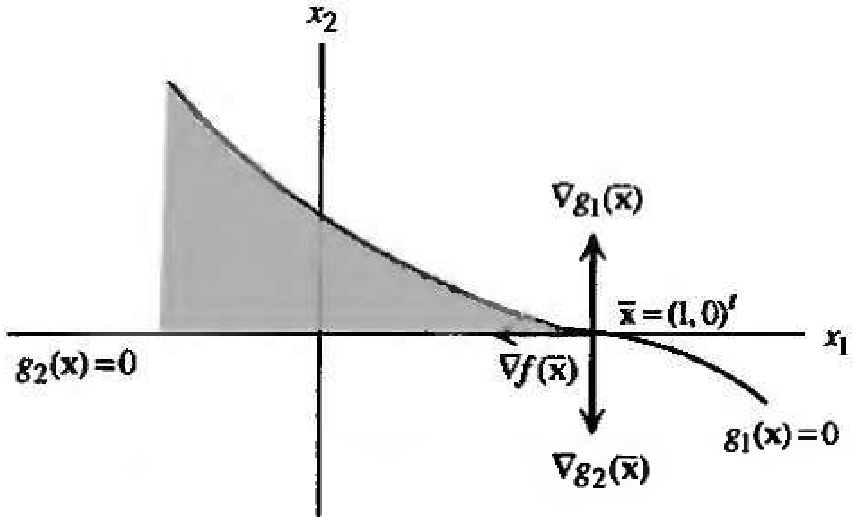


Figure 4.7 Feasible region in Example 4.2.10.

is true only if $u_0 = 0$. Thus, the Fritz John conditions hold true at $\bar{\mathbf{x}}$ by letting $u_0 = 0$ and $u_1 = u_2 = \alpha$, where α is any positive scalar.

4.2.11 Example

$$\begin{aligned} &\text{Minimize } -x_1 \\ &\text{subject to } x_1 + x_2 - 1 \le 0 \\ &\qquad\qquad\quad -x_2 \le 0. \end{aligned}$$

The feasible region is sketched in Figure 4.8, and the optimal point is $\bar{\mathbf{x}} = (1,0)^t$. Note that

$$\nabla f(\bar{\mathbf{x}}) = (-1,0)^t, \qquad \nabla g_1(\bar{\mathbf{x}}) = (1,1)^t, \qquad \nabla g_2(\bar{\mathbf{x}}) = (0,-1)^t,$$

and the Fritz John conditions hold true with $u_0 = u_1 = u_2 = \alpha$ for any positive scalar α.

As in the case of Theorem 4.2.5, there are points that satisfy the Fritz John conditions trivially. If a feasible point $\bar{\mathbf{x}}$ satisfies $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ or $\nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$ for some $i \in I$, clearly we can let the corresponding Lagrangian multiplier be any positive number, set all the other multipliers equal to zero, and satisfy the conditions of Theorem 4.2.8. The Fritz John conditions of Theorem 4.2.8 also hold true trivially at each feasible point for problems having equality constraints if each equality constraint is replaced by two equivalent inequalities. Specifically, if $g(\mathbf{x}) = 0$ is replaced by $g_1(\mathbf{x}) \equiv g(\mathbf{x}) \le 0$ and $g_2(\mathbf{x}) \equiv -g(\mathbf{x}) \le 0$, the Fritz John conditions are satisfied by taking $u_1 = u_2 = \alpha$ and setting all the other multipliers equal to zero, where α is any positive scalar.

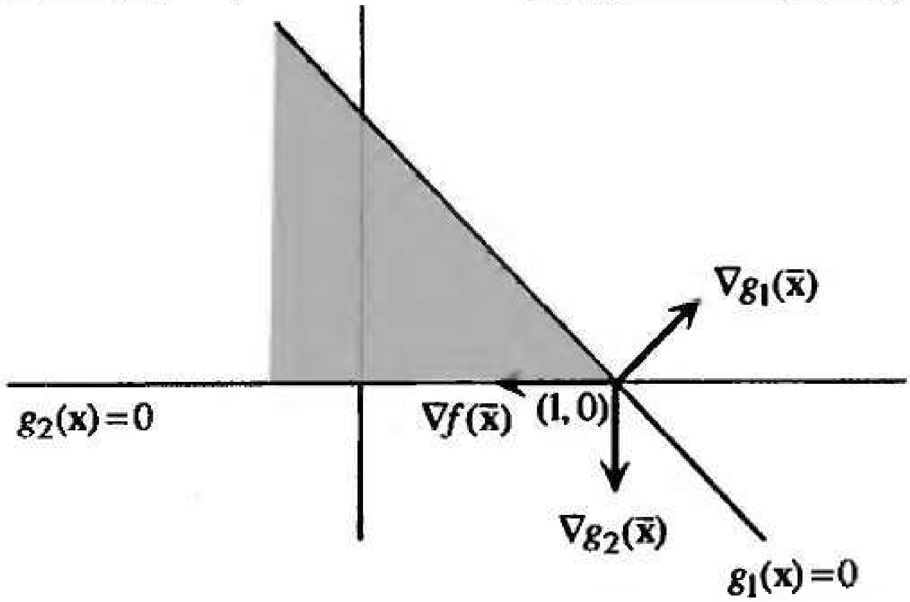


Figure 4.8 Feasible Region in Example 4.2.11.

In fact, given *any feasible solution* $\bar{x}$ to the problem of minimizing $f(x)$ subject to $x \in S$, we can add a redundant constraint to the problem to make $\bar{x}$ a FJ point! Specifically, we can add the constraint $\|x-\bar{x}\|^2 \geq 0$, which holds true for all $x \in R^n$. In particular, this constraint is binding at $\bar{x}$ and its gradient is also zero at $\bar{x}$. Consequently, we obtain $F_0 \cap G_0 = \emptyset$ at $\bar{x}$ since $G_0 = \emptyset$; so $\bar{x}$ is a FJ point!

This leads us to consider two issues. The first pertains to a set of conditions under which we can claim local optimality for a FJ point, and this is addressed in Theorem 4.2.12. The second consideration leads to the Karush–Kuhn–Tucker necessary optimality conditions, and this is addressed subsequently.

4.2.12 Theorem (Fritz John Sufficient Conditions)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$ and g_i: $R^n \to R$ for i = 1,..., m. Consider Problem P to minimize $f(x)$ subject to $x \in S$, and $g_i(x) \leq 0$ for i = 1,..., m. Let $\bar{x}$ be a FJ solution and denote $I = \{i\text{: } g_i(\bar{x}) = 0\}$. Define S as the relaxed feasible region for Problem P in which the nonbinding constraints are dropped.

a. If there exists an ε-neighborhood $N_\varepsilon(\bar{x})$, $\varepsilon > 0$, such that f is pseudoconvex over $N_\varepsilon(\bar{x}) \cap S$, and g_i, $i \in I$, are strictly pseudoconvex over $N_\varepsilon(\bar{x}) \cap S$, $\bar{x}$ is a local minimum for Problem P.
b. If f is pseudoconvex at $\bar{x}$, and if g_i, $i \in I$, are both strictly pseudoconvex and quasiconvex at $\bar{x}$, $\bar{x}$ is a global optimal solution for Problem P. In particular, if these generalized convexity assumptions

hold true only by restricting the domain of f to $N_\varepsilon(\bar{\mathbf{x}})$ for some $\varepsilon > 0$, $\bar{\mathbf{x}}$ is a local minimum for Problem P.

Proof

Suppose that the condition of Part a holds true. Since $\bar{\mathbf{x}}$ is a FJ point, we have equivalently by Gordan's theorem that $F_0 \cap G_0 = \varnothing$. By restricting attention to $S \cap N_\varepsilon(\bar{\mathbf{x}})$, we have, by closely following the proof to the converse statement of Theorem 4.2.5, that $\bar{\mathbf{x}}$ is a local minimum. This proves Part a.

Next, consider Part b. Again we have $F_0 \cap G_0 = \varnothing$. By restricting attention to S, we have $G_0 = D$ by Lemma 4.2.4; so we conclude that $F_0 \cap D = \varnothing$. Now let $\mathbf{x}$ be any feasible solution to the relaxed constraint set S. [In the case when the generalized convexity assumptions hold true over $N_\varepsilon(\bar{\mathbf{x}})$ alone, let $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$.] Since $g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}}) = 0$ for all $i \in I$, we have, by the quasiconvexity of g_i at $\bar{\mathbf{x}}$ for all $i \in I$, that

$$g_i[\mathbf{x} + \lambda(\mathbf{x} - \bar{\mathbf{x}})] = g_i[\lambda\mathbf{x} + (1-\lambda)\bar{\mathbf{x}}] \le \max\{g_i(\mathbf{x}), g_i(\bar{\mathbf{x}})\} = g_i(\bar{\mathbf{x}}) = 0$$

for all $0 \le \lambda \le 1$, and for each $i \in I$. This means that the direction $\mathbf{d} = (\mathbf{x} - \bar{\mathbf{x}}) \in D$. Because $F_0 \cap D = \varnothing$, we must therefore have $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} \ge 0$; that is, $\nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}) \ge 0$. By the pseudoconvexity of f at $\bar{\mathbf{x}}$, this in turn implies that $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$. Hence, $\bar{\mathbf{x}}$ is a global optimum over the relaxed set S [or over $S \cap N_\varepsilon(\bar{\mathbf{x}})$ in the second case]. And since it belongs to the original feasible region or to its intersection with $N_\varepsilon(\bar{\mathbf{x}})$, it is a global (or a local in the second case) minimum for Problem P. This completes the proof.

We remark here that as is evident from the analysis thus far, several variations of the assumptions in Theorem 4.2.12 are possible. We encourage the reader to explore this in Exercise 4.22.

Karush–Kuhn–Tucker Conditions

We have observed above that a point $\bar{\mathbf{x}}$ is a FJ point if and only if $F_0 \cap G_0 = \varnothing$. In particular, this condition holds true at any feasible solution $\bar{\mathbf{x}}$ at which $G_0 = \varnothing$, regardless of the objective function. For example, if the feasible region has no interior in the immediate vicinity of $\bar{\mathbf{x}}$, or if the gradient of some binding constraint (which might even be redundant) vanishes, $G_0 = \varnothing$. Generally speaking, by Gordan's theorem, $G_0 = \varnothing$ if and only if the gradients of the binding constraints can be made to cancel out using nonnegative, nonzero linear combinations, and whenever this case occurs, $\bar{\mathbf{x}}$ will be a FJ point. More disturbingly, it follows that FJ points can be nonoptimal even for the well-behaved and important class of linear programming (LP) problems. Figure 4.9 illustrates this situation.

Motivated by this observation, we are led to the KKT conditions described next that encompass FJ points for which there exist Lagrange multipliers such that $u_0 > 0$ and hence force the objective function gradient to play a role in the optimality conditions. These conditions were derived independently by Karush [1939] and by Kuhn and Tucker [1951], and are precisely the FJ conditions with the added requirement that $u_0 > 0$. Note that when $u_0 > 0$, by scaling the dual feasibility conditions, if necessary, we can assume without loss of generality that $u_0 \equiv 1$. Hence, in Example 4.2.9, taking $u_0 = 1$ in the FJ conditions, we obtain (u_0, u_1, u_2) = (1, 1/3, 2/3) as the Lagrange multipliers corresponding to the optimal solution. Moreover, in Figure 4.9, the only FJ point that is also a KKT point is the optimal solution $\bar{\mathbf{x}}$. In fact, as we shall see later, the KKT conditions are both necessary and sufficient for optimality for linear programming problems. Example 4.2.11 gives another illustration of a linear programming problem.

Also, note from the above discussion that if $G_0 \neq \varnothing$ at a local minimum $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ must indeed be a KKT point; that is, it must be a FJ point with $u_0 > 0$. This follows because by Gordan's theorem, if $G_0 \neq \varnothing$, no solution exists to FJ's dual feasibility conditions with $u_0 = 0$. Hence, $G_0 \neq \varnothing$ is a *sufficient* condition placed on the behavior of the constraints to ensure that a local minimum $\bar{\mathbf{x}}$ is a KKT point. Of course, it need not necessarily hold true whenever a local minimum $\bar{\mathbf{x}}$ turns out to be a KKT point, as in Figure 4.9, for example. Such a condition is known as a *constraint qualification* (CQ). Several conditions of this kind are

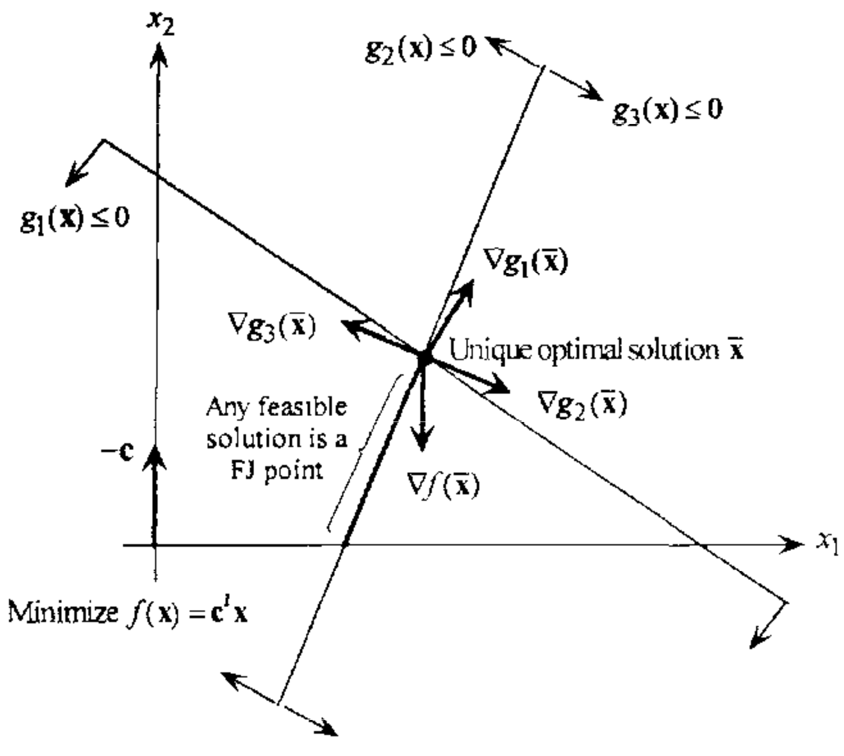


Figure 4.9 FJ conditions are not sufficient for optimality for LP problems.

discussed in more detail later and in Chapter 5. Note that the importance of constraint qualifications is to guarantee that by examining only KKT points, we do not lose out on local minima and hence, possibly, global optimal solutions. This can certainly occur, as is evident from Figure 4.7 of Example 4.2.10, where u_0 is necessarily zero in the FJ conditions for the optimal solution.

In Theorem 4.2.13, by imposing the constraint qualification that the gradient vectors of the binding constaints are linearly independent, we obtain the KKT conditions. Note that if the gradients of the binding constraints are linearly independent, certainly they cannot be canceled by using nonzero, nonnegative, linear combinations; and hence, this implies by Gordan's theorem that $G_0 \neq \varnothing$. Therefore, the linear independence constraint qualification implies the constraint qualification that $G_0 \neq \varnothing$; and hence, as above, it implies that a local minimum $\bar{\mathbf{x}}$ satisfies the KKT conditions. This is formalized below.

4.2.13 Theorem (Karush–Kuhn–Tucker Necessary Conditions)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$ and g_i: $R^n \to R$ for $i =$ 1,..., m. Consider the Problem P to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \leq$ 0 for $i = 1,..., m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that g_i for $i \notin I$ are continuous at $\bar{\mathbf{x}}$. Furthermore, suppose that $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ are linearly independent. If $\bar{\mathbf{x}}$ solves Problem P locally, there exist scalars u_i for $i \in I$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i \geq 0 \qquad \text{for } i \in I.$$

In addition to the above assumptions, if g_i for each $i \notin I$ is also differentiable at $\bar{\mathbf{x}}$, the foregoing conditions can be written in the following equivalent form:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots,m.$$

Proof

By Theorem 4.2.8 there exist scalars u_0 and $\hat{u}_i$ for $i \in I$, not all equal to zero, such that

$$\begin{aligned} u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} \hat{u}_i \nabla g_i(\bar{\mathbf{x}}) &= \mathbf{0} \\ u_0, \hat{u}_i &\geq 0 \qquad \text{for } i \in I. \end{aligned} \tag{4.10}$$

Note that $u_0 > 0$ because (4.10) would contradict the assumption of linear independence of $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ if $u_0 = 0$. The first part of the theorem follows by letting $u_i = \hat{u}_i/u_0$ for each $i \in I$. The equivalent form of the necessary conditions follows by letting $u_i = 0$ for $i \notin I$. This completes the proof.

As in the Fritz John conditions, the scalars u_i are called the *Lagrangian*, or *Lagrange, multipliers*. The requirement that $\bar{\mathbf{x}}$ be feasible to Problem P is called the *primal feasibility* (PF) *condition*, whereas the condition that $\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = 0$, $u_i \geq 0$ for $i = 1,\ldots, m$, is referred to as the *dual feasibility* (DF) *condition*. The restriction $u_i g_i(\bar{\mathbf{x}}) = 0$ for each $i = 1,\ldots,m$ is called the *complementary slackness* (CS) *condition*. Together, these PF, DF, and CS conditions are called the *Karush–Kuhn–Tucker conditions*. Any point $\bar{\mathbf{x}}$ for which there exist Lagrangian (or Lagrange) multipliers $\bar{\mathbf{u}}$ such that $(\bar{\mathbf{x}}, \bar{\mathbf{u}})$ satisfies the KKT conditions is called a *KKT point*. Note that if the gradients $\nabla g_i(\bar{\mathbf{x}})$, $i \in I$, are linearly independent, then by the DF and CS conditions, the associated Lagrange multipliers are determined uniquely at the KKT point $\bar{\mathbf{x}}$.

The KKT conditions can alternatively be written in vector form as follows, in addition to the PF requirement:

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) + \nabla \mathbf{g}(\bar{\mathbf{x}})^t \mathbf{u} &= 0 \\ \mathbf{u}^t \mathbf{g}(\bar{\mathbf{x}}) &= 0 \\ \mathbf{u} &\geq 0. \end{aligned}$$

Here $\nabla \mathbf{g}(\bar{\mathbf{x}})^t$ is an $n \times m$ matrix whose *i*th column is $\nabla g_i(\bar{\mathbf{x}})$ (it is the transpose of the *Jacobian* of $\mathbf{g}$ at $\bar{\mathbf{x}}$), and $\mathbf{u}$ is an m-vector denoting the Lagrangian multipliers.

Now, consider Examples 4.2.9, 4.2.10, and 4.2.11. In Example 4.2.9, at $\bar{\mathbf{x}} = (2,1)^t$, the reader may verify that $u_1 = 1/3$, $u_2 = 2/3$, and $u_3 = u_4 = 0$ will satisfy the KKT conditions. Example 4.2.10 does not satisfy the assumptions of Theorem 4.2.13 at $\bar{\mathbf{x}} = (1,0)^t$, since $\nabla g_1(\bar{\mathbf{x}})$ and $\nabla g_2(\bar{\mathbf{x}})$ are linearly dependent. In fact, in this example, we saw that u_0 is necessarily zero in the FJ conditions. In Example 4.2.11, $\bar{\mathbf{x}} = (1,0)^t$ and $u_1 = u_2 = 1$ satisfy the KKT conditions.

4.2.14 Example (Linear Programming Problems)

Consider the linear programming Problem P: Minimize $\{\mathbf{c}^t\mathbf{x}: \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $\mathbf{A}$ is $m \times n$ and the other vectors are conformable. Writing the constraints as $-\mathbf{A}\mathbf{x} \leq -\mathbf{b}$, $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, and $-\mathbf{x} \leq \mathbf{0}$, and denoting the Lagrange multiplier vectors with respect to these three sets as $\mathbf{y}^+$, $\mathbf{y}^-$, and $\mathbf{v}$, respectively, the KKT conditions are as follows:

$$\text{PF}: \quad \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

$$\text{DF}: \quad -\mathbf{A}^t\mathbf{y}^+ + \mathbf{A}^t\mathbf{y}^- - \mathbf{v} = -\mathbf{c}, (\mathbf{y}^+, \mathbf{y}^-, \mathbf{v}) \geq \mathbf{0}$$

$$\text{CS}: \quad (\mathbf{b} - \mathbf{Ax})^t\mathbf{y}^+ = 0, \qquad (\mathbf{Ax} - \mathbf{b})^t\mathbf{y}^- = 0, \qquad -\mathbf{x}^t\mathbf{v} = 0.$$

Denoting $\mathbf{y} = \mathbf{y}^+ - \mathbf{y}^-$ as the difference of the two nonnegative variable vectors $\mathbf{y}^+$ and $\mathbf{y}^-$, we can equivalently write the KKT conditions as follows, noting the use of the PF and DF conditions in simplifying the CS conditions:

$$\text{PF}: \quad \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

$$\text{DF}: \quad \mathbf{A}^t\mathbf{y} + \mathbf{v} = \mathbf{c}, \mathbf{v} \geq 0, \quad (\mathbf{y} \text{ unrestricted})$$

$$\text{CS}: \quad x_j v_j = 0 \qquad \text{for } j = 1, \ldots, n.$$

Hence, from Theorem 2.7.3 and its Corollary 3, observe that $\bar{\mathbf{x}}$ is KKT solution with associated Lagrange multipliers $(\bar{\mathbf{y}}, \bar{\mathbf{v}})$ if and only if $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are, respectively, optimal to the primal and dual linear programs P and D, where D: Maximize $\{\mathbf{b}^t\mathbf{y} : \mathbf{A}^t\mathbf{y} \leq \mathbf{c}\}$. In particular, observe that the DF restriction in the KKT conditions is precisely the feasibility condition to the dual D: hence, the name. This example therefore establishes that for linear programming problems, the KKT conditions are both necessary and sufficient for optimality to the primal and dual problems.

Geometric Interpretation of the Karush–Kuhn–Tucker Conditions: Connections with Linear Programming Approximations

Note that any vector of the form $\sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}})$, where $u_i \geq 0$ for $i \in I$, belongs to the cone spanned by the gradients of the binding constraints. The KKT dual feasibility conditions $-\nabla f(\bar{\mathbf{x}}) = \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}})$ and $u_i \geq 0$ for $i \in I$ can then be interpreted as $-\nabla f(\bar{\mathbf{x}})$, belonging to the cone spanned by the gradients to the binding constraints at a given feasible solution $\bar{\mathbf{x}}$.

Figure 4.10 illustrates this concept for two points $\mathbf{x}_1$ and $\mathbf{x}_2$. Note that $-\nabla f(\mathbf{x}_1)$ belongs to the cone spanned by the gradients of the binding constraints at $\mathbf{x}_1$ and hence $\mathbf{x}_1$ is a KKT point; that is, $\mathbf{x}_1$ satisfies the KKT conditions. On the other hand, $-\nabla f(\mathbf{x}_2)$ lies outside the cone spanned by the gradients of the binding constraints at $\mathbf{x}_2$ and thus contradicts the KKT conditions.

Similarly, in Figures 4.6 and 4.8, for $\bar{\mathbf{x}} = (2, 1)^t$ and $\bar{\mathbf{x}} = (1, 0)^t$, respectively, $-\nabla f(\bar{\mathbf{x}})$ lies in the cone spanned by the gradients of the binding constraints at $\bar{\mathbf{x}}$. On the other hand, in Figure 4.7, for $\bar{\mathbf{x}} = (1, 0)^t$, $-\nabla f(\bar{\mathbf{x}})$ lies outside the cone spanned by the gradients of the binding constraints at $\bar{\mathbf{x}}$.

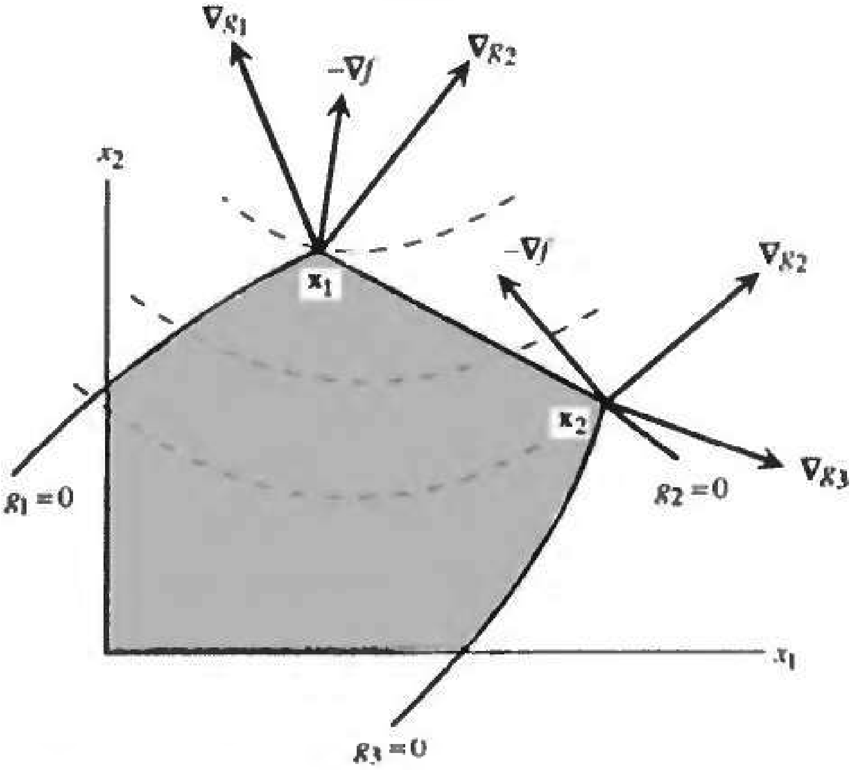


Figure 4.10 Geometric illustration of the KKT conditions.

We now provide a key insight into the KKT conditions via linear programming duality and Farkas's lemma as expounded in Theorem 2.7.3 and its Corollary 2. The following result asserts that a feasible solution $\bar{\mathbf{x}}$ is a KKT point if and only if it happens to be an optimal to the linear program obtained by replacing the objective and the constraints by their first-order approximations at $\bar{\mathbf{x}}$. (This is referred to as the *first-order linear programming approximation* to the problem at $\bar{\mathbf{x}}$.) Not only does this provide a useful conceptual characterization of KKT points and an insight into its value and interpretation, but it affords a useful construct in deriving algorithms that are designed to converge to a KKT solution.

4.2.15 Theorem (KKT Conditions and First-Order LP Approximations)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$ and g_i: $R^n \to R$, $i = 1,\ldots, m$ be differentiable functions. Consider Problem P, to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S = \{\mathbf{x} \in X: g_i(\mathbf{x}) \le 0, i = 1,\ldots, m\}$. Let $\bar{\mathbf{x}}$ be a feasible solution, and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Define $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$ and $G_0' = \{\mathbf{d} \neq \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0$, for each $i \in I\}$ as before, and let $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0$ for each $i \in I\} = G_0' \cup \{\mathbf{0}\}$. Then $\bar{\mathbf{x}}$ is a KKT solution if and only if $F_0 \cap G' = \varnothing$, which is equivalent to $F_0 \cap G_0' = \varnothing$. Furthermore, consider the *first-order linear programming approximation* to Problem P:

$$\text{LP}(\bar{\mathbf{x}}): \text{ Minimize}\{f(\bar{\mathbf{x}})+\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}): g_i(\bar{\mathbf{x}})+\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \le 0$$
$$\text{for } i=1,\ldots,m\}.$$

Then, $\bar{\mathbf{x}}$ is a KKT solution if and only if $\bar{\mathbf{x}}$ solves LP($\bar{\mathbf{x}}$).

Proof

The feasible solution $\bar{\mathbf{x}}$ is a KKT point if and only if there exists a solution (u_i, $i \in I$) to the system $\sum_{i\in I} u_i \nabla g_i(\bar{\mathbf{x}}) = -\nabla f(\bar{\mathbf{x}})$ and $u_i \ge 0$ for $i \in I$. By Farkas's lemma (see, e.g., Corollary 2 to Theorem 2.7.3), this holds true if and only if there does not exist a solution to the system $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0$ for $i \in I$ and $\nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0$. Hence, $\bar{\mathbf{x}}$ is a KKT point if and only if $F_0 \cap G' = \varnothing$. Clearly, we also have that this holds true if and only if $F_0 \cap G_0' = \varnothing$.

Now consider the first-order linear programming approximation LP($\bar{\mathbf{x}}$) given in the theorem. The solution $\bar{\mathbf{x}}$ is obviously feasible to LP. Ignoring the constant terms in the objective function and writing LP($\bar{\mathbf{x}}$) in the form of Problem D of Theorem 2.7.3, we get that, equivalently, LP($\bar{\mathbf{x}}$): Maximize $\{-\nabla f(\bar{\mathbf{x}})^t \mathbf{x} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{x} \le \nabla g_i(\bar{\mathbf{x}})^t \bar{\mathbf{x}} - g_i(\bar{\mathbf{x}}) \text{ for } i=1,\ldots,m\}$. The dual to this problem, denoted DLP($\bar{\mathbf{x}}$), is to

$$\text{Minimize} \quad \sum_{i=1}^{m} u_i[\nabla g_i(\bar{\mathbf{x}})^t \bar{\mathbf{x}} - g_i(\bar{\mathbf{x}})]$$
$$\text{subject to} \quad \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = -\nabla f(\bar{\mathbf{x}}), \quad u_i \ge 0 \quad \text{for } i=1,\ldots,m.$$

Hence, by Corollary 3 to Theorem 2.7.3, we deduce that $\bar{\mathbf{x}}$ is an optimal solution to LP($\bar{\mathbf{x}}$) if and only if there exists a solution $\bar{\mathbf{u}}$ feasible to DLP($\bar{\mathbf{x}}$) that also satisfies the complementary slackness condition $\bar{u}_i[\nabla g_i(\bar{\mathbf{x}})^t \bar{\mathbf{x}} - \nabla g_i(\bar{\mathbf{x}})^t \bar{\mathbf{x}} + g_i(\bar{\mathbf{x}})] = \bar{u}_i g_i(\bar{\mathbf{x}}) = 0$ for $i = 1,\ldots, m$. But this is precisely the KKT conditions. Hence, $\bar{\mathbf{x}}$ is optimal to LP($\bar{\mathbf{x}}$) if and only if $\bar{\mathbf{x}}$ is a KKT solution for P, and this completes the proof.

To illustrate, observe that in Figure 4.6 of Example 4.2.9, if we replace $g_i(\mathbf{x}) \le 0$ by its tangential first-order approximation at the point (2, 1) and the objective function by the linear objective of minimizing $\nabla f(\bar{\mathbf{x}})^t \mathbf{x}$, the given point (2, 1) is optimal to the resulting linear programming problem and hence is a KKT solution. On the other hand, in Figure 4.7 of Example 4.2.10, the feasible region for the linear programming approximation at $\bar{\mathbf{x}} = (1, 0)^t$ is the entire x_1-axis. Clearly, then, the point (1, 0) is not optimal to the underlying linear program LP($\bar{\mathbf{x}}$) of minimizing $\nabla f(\bar{\mathbf{x}})^t \mathbf{x}$ over this region, and thus the point (1, 0) is not a KKT point. Hence, the KKT conditions, being oblivious to the nonlinear behavior of the constraint $g_1(\mathbf{x}) \le 0$ about the point $\bar{\mathbf{x}}$ other than its first-order approximation, fail to recognize the optimality of this solution for the original nonlinear problem.

Theorem 4.2.16 shows that under convexity assumptions, the KKT conditions are also sufficient for (local) optimality.

4.2.16 Theorem (Karush–Kuhn–Tucker Sufficient Conditions)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$ and g_i: $R^n \to R$ for i = 1,..., m. Consider Problem P, to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \le 0$ for i = 1,..., m. Let $\bar{\mathbf{x}}$ be a KKT solution, and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Define S as the relaxed feasible region for Problem P in which the constraints that are not binding at $\bar{\mathbf{x}}$ are dropped. Then:

a. If there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ about $\bar{\mathbf{x}}$, $\varepsilon > 0$ such that f is pseudoconvex over $N_\varepsilon(\bar{\mathbf{x}}) \cap S$ and g_i, $i \in I$, are differentiable at $\bar{\mathbf{x}}$ and are quasiconvex over $N_\varepsilon(\bar{\mathbf{x}}) \cap S$, $\bar{\mathbf{x}}$ is a local minimum for Problem P.
b. If f is pseudoconvex at $\bar{\mathbf{x}}$, and if g_i, $i \in I$, are differentiable and quasiconvex at $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is a global optimal solution to Problem P. In particular, if this assumption holds true with the domain of the feasible restriction to $N_\varepsilon(\bar{\mathbf{x}})$, for some $\varepsilon > 0$, $\bar{\mathbf{x}}$ is a local minimum for P.

Proof

First, consider Part a. Since $\bar{\mathbf{x}}$ is a KKT point, we have, equivalently, by Theorem 4.2.15 that $F_0 \cap G_0' = \varnothing$. From (4.7) this means that $F_0 \cap D = \varnothing$. Since g_i, $i \in I$, are quasiconvex over $N_\varepsilon(\bar{\mathbf{x}}) \cap S$, we have that $N_\varepsilon(\bar{\mathbf{x}}) \cap S$ is a convex set. By restricting attention to $N_\varepsilon(\bar{\mathbf{x}}) \cap S$, we therefore have the condition of the converse statement of Theorem 4.2.2 holding true; so $\bar{\mathbf{x}}$ is a minimum over $N_\varepsilon(\bar{\mathbf{x}}) \cap S$. Hence, $\bar{\mathbf{x}}$ is a local minimum for the more restricted original Problem P. This proves Part a.

Next, consider Part b. Let $\mathbf{x}$ be any feasible solution to Problem P. [In case the generalized convexity definitions are restricted to $N_\varepsilon(\bar{\mathbf{x}})$, let $\mathbf{x}$ be any feasible solution to P that lies within $N_\varepsilon(\bar{\mathbf{x}})$.] Then, for $i \in I$, $g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}})$, since $g_i(\mathbf{x}) \le 0$ and $g_i(\bar{\mathbf{x}}) = 0$. By the quasiconvexity of g_i at $\bar{\mathbf{x}}$, it follows that

$$g_i[\bar{\mathbf{x}} + \lambda(\mathbf{x} - \bar{\mathbf{x}})] = g_i[\lambda\mathbf{x} + (1-\lambda)\bar{\mathbf{x}}] \le \max\{g_i(\mathbf{x}), g_i(\bar{\mathbf{x}})\} = g_i(\bar{\mathbf{x}})$$

for all $\lambda \in (0, 1)$. This implies that g_i does not increase by moving from $\bar{\mathbf{x}}$ along the direction $\mathbf{x} - \bar{\mathbf{x}}$. Thus, by Theorem 4.1.2 we must have $\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \le 0$.

Multiplying this by the Lagrange multiplier u_i corresponding to the KKT point $\bar{\mathbf{x}}$, and summing over I, we get $[\sum_{i\in I} u_i \nabla g_i(\bar{\mathbf{x}})^t](\mathbf{x}-\bar{\mathbf{x}}) \le 0$. But since $\nabla f(\bar{\mathbf{x}})$ + $\sum_{i\in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$, it follows that $\nabla f(\bar{\mathbf{x}})^t(\mathbf{x}-\bar{\mathbf{x}}) \ge 0$. Then, by the pseudoconvexity of f at $\bar{\mathbf{x}}$, we must have $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$, and the proof is complete.

Needless to say, if f and g_i are convex at $\bar{\mathbf{x}}$, and hence both pseudoconvex and quasiconvex at $\bar{\mathbf{x}}$, the KKT conditions are sufficient. Also, if convexity at a point is replaced by the stronger requirement of global convexity, the KKT conditions are also sufficient for global optimality. (We ask the reader to explore other variations of this result in Exercises 4.22 and 4.50.)

There is one important point to note in regard to KKT conditions that is often a source of error. Namely, despite the usually well-behaved nature of convex programming problems and the sufficiency of KKT conditions under convexity assumptions, the KKT conditions are *not* necessary for optimality for convex programming problems. Figure 4.11 illustrates this situation for the convex programming problem given:

$$\begin{aligned} &\text{Minimize } x_1 \\ &\text{subject to } (x_1-1)^2+(x_2-1)^2 \le 1 \\ &\qquad\qquad\; (x_1-1)^2+(x_2+1)^2 \le 1. \end{aligned}$$

The only feasible solution $\bar{\mathbf{x}} = (1,0)^t$ is naturally optimal. However, this is not a KKT point. Note in connection with Theorem 4.2.15 that the first-order linear programming approximation at $\bar{\mathbf{x}}$ is unbounded. However, as we shall see in Chapter 5, if there exists an interior point feasible solution to the set of constraints that are binding at an optimum $\bar{\mathbf{x}}$ to a convex programming problem, $\bar{\mathbf{x}}$ is indeed a KKT point and is therefore captured by the KKT conditions.

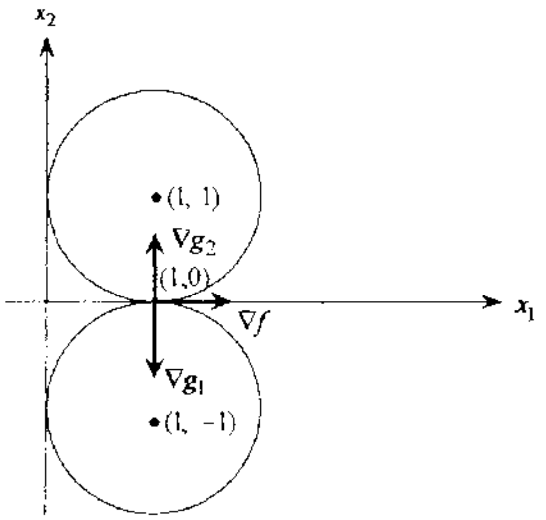


Figure 4.11 KKT conditions are not necessary for convex programming problems.

4.3 Problems Having Inequality and Equality Constraints

In this section we generalize the optimality conditions of the preceding section to handle inequality constraints as well as equality constraints. Consider the following nonlinear programming Problem P:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 \quad \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. \end{aligned}$$

As a natural extension of Theorem 4.2.5, in Theorem 4.3.1 we show that if $\overline{\mathbf{x}}$ is a local optimal solution to Problem P, either the gradients of the equality constraints are linearly dependent at $\overline{\mathbf{x}}$, or else, $F_0 \cap G_0 \cap H_0 = \varnothing$, where $H_0 = \{\mathbf{d} : \nabla h_i(\overline{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\ldots,\ell\}$. A reader with only a casual interest in the derivation of optimality conditions may skip the proof of Theorem 4.3.1, since it involves the more advanced concepts of solving a system of differential equations.

4.3.1 Theorem

Let X be a nonempty open set in R^n. Let f: $R^n \to R$, g_i: $R^n \to R$ for $i = 1,\ldots, m$, and h_i: $R^n \to R$ for $i = 1,\ldots,\ell$. Consider the Problem P given below:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 \quad \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. \end{aligned}$$

Suppose that $\overline{\mathbf{x}}$ is a local optimal solution, and let $I = \{i\colon g_i(\overline{\mathbf{x}}) = 0\}$. Furthermore, suppose that each g_i for $i \notin I$ is continuous at $\overline{\mathbf{x}}$, that f and g_i for $i \in I$ are differentiable at $\overline{\mathbf{x}}$ and that each h_i for $i = 1,\ldots,\ell$ is continuously differentiable at $\overline{\mathbf{x}}$. If $\nabla h_i(\overline{\mathbf{x}})$ for $i = 1,\ldots,\ell$ are linearly independent, $F_0 \cap G_0 \cap H_0 = \varnothing$, where

$$\begin{aligned} F_0 &= \{\mathbf{d} : \nabla f(\overline{\mathbf{x}})^t \mathbf{d} < 0\} \\ G_0 &= \{\mathbf{d} : \nabla g_i(\overline{\mathbf{x}})^t \mathbf{d} < 0 \text{ for } i \in I\} \\ H_0 &= \{\mathbf{d} : \nabla h_i(\overline{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\ldots,\ell\}. \end{aligned}$$

Conversely, suppose that $F_0 \cap G_0 \cap H_0 = \varnothing$. If f is pseudoconvex at $\overline{\mathbf{x}}$, g_i for $i \in I$ are strictly pseudoconvex over some ε-neighborhood of $\overline{\mathbf{x}}$; and if h_i for $i = 1,\ldots,\ell$ are affine, $\overline{\mathbf{x}}$ is a local optimal solution.

Proof

Consider the first part of the theorem. By contradiction, suppose that there exists a vector $\mathbf{y} \in F_0 \cap G_0 \cap H_0$; that is, $\nabla f(\bar{\mathbf{x}})^t \mathbf{y} < 0$, $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{y} < 0$ for each $i \in I$, and $\nabla \mathrm{h}(\bar{\mathbf{x}})\mathbf{y} = \mathbf{0}$, where $\nabla \mathbf{h}(\bar{\mathbf{x}})$ is the $\ell \times n$ Jacobian matrix whose ith row is $\nabla h_i(\bar{\mathbf{x}})^t$. Let us now construct a feasible arc from $\bar{\mathbf{x}}$ obtained by projecting points along $\mathbf{y}$ from $\bar{\mathbf{x}}$ onto the equality-constraint surface. For $\lambda \geq 0$, define $\alpha: R \to R^n$ by the following differential equation and boundary condition:

$$\frac{d\alpha(\lambda)}{d\lambda} = \mathbf{P}(\lambda)\mathbf{y} \qquad \text{and} \qquad \alpha(0) = \bar{\mathbf{x}}, \tag{4.11}$$

where $\mathbf{P}(\lambda)$is the matrix that projects any vector in the null space of $\nabla \mathrm{h}[\alpha(\lambda)]$. For λ sufficiently small, (4.11) is well defined and solvable because $\nabla \mathbf{h}(\bar{\mathbf{x}})$ has full rank and $\mathbf{h}$ is continuously differentiable at $\bar{\mathbf{x}}$, so that $\mathbf{P}$ is continuous in λ. Obviously, $\alpha(\lambda) \to \bar{\mathbf{x}}$ as $\lambda \to 0^+$.

We now show that for $\lambda > 0$ and sufficiently small, $\alpha(\lambda)$ is feasible and $f[\alpha(\lambda)] < f(\bar{\mathbf{x}})$, thus contradicting local optimality of $\bar{\mathbf{x}}$. By the chain rule of differentiation and from (4.11), we get

$$\frac{d}{d\lambda} g_i[\alpha(\lambda)] = \nabla g_i[\alpha(\lambda)]^t \mathbf{P}(\lambda)\mathbf{y} \tag{4.12}$$

for each $i \in I$. In particular, $\mathbf{y}$ is in the null space of $\nabla \mathbf{h}(\bar{\mathbf{x}})$, so for $\lambda = 0$, we have $\mathbf{P}(0)\mathbf{y} = \mathrm{y}$. Hence, from (4.12) and the fact that $\nabla g_i(\bar{\mathbf{x}})^t \mathrm{y} < 0$, we get

$$\frac{d}{d\lambda} g_i[\alpha(0)] = \nabla g_i(\bar{\mathbf{x}})^t \mathrm{y} < 0 \tag{4.13}$$

for $i \in I$. This implies further that $g_i[\alpha(\lambda)] < 0$ for $\lambda > 0$ and sufficiently small. For $i \notin I$, $g_i(\bar{\mathbf{x}}) < 0$, and g_i is continuous at $\bar{\mathbf{x}}$, and thus $g_i[\alpha(\lambda)] < 0$ for λ sufficiently small. Also, since X is open, $\alpha(\lambda) \in X$ for λ sufficiently small. To show feasibility of $\alpha(\lambda)$, we only need to show that $h_i[\alpha(\lambda)] = 0$ for λ sufficiently small. By the mean value theorem, we have

$$\begin{aligned} h_i[\alpha(\lambda)] &= h_i[\alpha(0)] + \lambda \frac{d}{d\lambda} h_i[\alpha(\mu)] \\ &= \lambda \frac{d}{d\lambda} h_i[\alpha(\mu)] \end{aligned} \tag{4.14}$$

for some $\mu \in (0, \lambda)$. But by the chain rule of differentiation and similar to (4.12), we get

$$\frac{d}{d\lambda}h_i[\alpha(\mu)] = \nabla h_i[\alpha(\mu)]^t \mathbf{P}(\mu)\mathbf{y}.$$

By construction, $\mathbf{P}(\mu)\mathbf{y}$ is in the null space of $\nabla h_i[\alpha(\mu)]$ and, hence, from the above equation, we get $(d/d\lambda)h_i[\alpha(\mu)] = 0$. Substituting in (4.14), it follows that $h_i[\alpha(\lambda)] = 0$. Since this is true for each i, it follows that $\alpha(\lambda)$ is a feasible solution to Problem P for each $\lambda > 0$ and is sufficiently small. By an argument similar to that leading to (4.13), we get

$$\frac{d}{d\lambda}f[\alpha(0)] = \nabla f(\bar{\mathbf{x}})^t \mathbf{y} < 0$$

and hence $f[\alpha(\lambda)] < f(\bar{\mathbf{x}})$ for $\lambda > 0$ and sufficiently small. This contradicts the local optimality of $\bar{\mathbf{x}}$. Hence, $F_0 \cap G_0 \cap H_0 = \varnothing$.

Conversely, suppose that $F_0 \cap G_0 \cap H_0 = \varnothing$ and that the assumptions of the converse statement to the theorem hold true. Since h_i, $i = 1,\ldots,\ell$, are affine, we have that $\mathbf{d}$ is a feasible direction for the equality constraints if and only if $\mathbf{d} \in H_0$. Using Lemma 4.2.4, it is readily verified that since g_i, $i \in I$, are strictly pseudoconvex over $N_\varepsilon(\bar{\mathbf{x}})$ for some $\varepsilon > 0$, we have that $D = G_0 \cap H_0$, where D is the set of feasible directions at $\bar{\mathbf{x}}$ defined for the set $S = \{\mathbf{x} : g_i(\mathbf{x}) \le 0 \text{ for } i \in I,\ h_i(\mathbf{x}) = 0 \text{ for } i = 1,\ldots,\ell\}$. Hence, we have that $F_0 \cap D = \varnothing$. Moreover, by our assumptions, we know that $S \cap N_\varepsilon(\bar{\mathbf{x}})$ is a convex set and that f is pseudoconvex at $\bar{\mathbf{x}}$. Hence, by the converse to Theorem 4.2.2, $\bar{\mathbf{x}}$ is a minimum over $S \cap N_\varepsilon(\bar{\mathbf{x}})$. Therefore, $\bar{\mathbf{x}}$ is a local minimum for the more restricted original problem as well, and this completes the proof.

Fritz John Conditions

We now express the geometric optimality condition $F_0 \cap G_0 \cap H_0 = \varnothing$ in a more usable algebraic form. This is done in Theorem 4.3.2, which is a generalization of the Fritz John conditions of Theorem 4.2.6.

4.3.2 Theorem (Fritz John Necessary Conditions)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$, g_i: $R^n \to R$ for $i = 1,\ldots, m$, and h_i: $R^n \to R$ for $i = 1,\ldots, \ell$. Consider Problem P defined below:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \qquad \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 \qquad \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. \end{aligned}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, suppose that g_i for each $i \in I$ is continuous at $\bar{\mathbf{x}}$, that f_i and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$ and that h_i for each $i = 1,..., \ell$ is continuously differentiable at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ solves Problem P locally, there exist scalars u_0, u_i for $i \in I$, and v_i for $i = 1,..., \ell$ such that

$$\begin{aligned} u_0\nabla f(\bar{\mathbf{x}}) + \sum_{i\in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) &= \mathbf{0} \\ u_0, u_i &\geq 0 \qquad \text{for } i \in I \\ (u_0, \mathbf{u}_I, \mathbf{v}) &\neq (0,\mathbf{0},\mathbf{0}), \end{aligned}$$

where $\mathbf{u}_I$ is the vector whose components are u_i for $i \in I$ and $\mathbf{v} = (v_1,...,v_\ell)^t$. Furthermore, if each g_i for $i \notin I$ is also differentiable at $\bar{\mathbf{x}}$, the Fritz John conditions can be written in the following equivalent form, where $\mathbf{u} = (u_1,...,u_m)^t$ and $\mathbf{v} = (v_1,..., v_\ell)^t$:

$$\begin{aligned} u_0\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) &= \mathbf{0} \\ u_i g_i(\bar{\mathbf{x}}) &= 0 \qquad \text{for } i = 1,..., m \\ u_0, u_i &\geq 0 \qquad \text{for } i = 1,..., m \\ (u_0, \mathbf{u}, \mathbf{v}) &\neq (0,\mathbf{0},\mathbf{0}). \end{aligned}$$

Proof

If $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,...,\ell$ are linearly dependent, we can find scalars $v_1,..., v_\ell$, not all zero, such that $\sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$. Letting u_0, u_i for $i \in I$ be equal to zero, the conditions of the first part of the theorem hold trivially.

Now, suppose that $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,...,\ell$ are linearly independent. Let $\mathbf{A}_1$ be the matrix whose rows are $\nabla f(\bar{\mathbf{x}})^t$ and $\nabla g_i(\bar{\mathbf{x}})^t$ for $i \in I$, and let $\mathbf{A}_2$ be the matrix whose rows are $\nabla h_i(\bar{\mathbf{x}})^t$ for $i = 1,...,\ell$. Then, from Theorem 4.3.1, the local optimality of $\bar{\mathbf{x}}$ implies that the system

$$\mathbf{A}_1\mathbf{d} < \mathbf{0}, \qquad \mathbf{A}_2\mathbf{d} = \mathbf{0}$$

is inconsistent. Now, consider the following two sets:

$$\begin{aligned} S_1 &= \{(\mathbf{z}_1, \mathbf{z}_2) : \mathbf{z}_1 = \mathbf{A}_1\mathbf{d},\ \mathbf{z}_2 = \mathbf{A}_2\mathbf{d}\} \\ S_2 &= \{(\mathbf{z}_1, \mathbf{z}_2) : \mathbf{z}_1 < \mathbf{0},\ \mathbf{z}_2 = \mathbf{0}\}. \end{aligned}$$

Note that S_1 and S_2 are nonempty convex sets such that $S_1 \cap S_2 = \varnothing$. Then by Theorem 2.4.8 there exists a nonzero vector $\mathbf{p}^t = (\mathbf{p}_1^t, \mathbf{p}_2^t)$ such that

$$\mathbf{p}_1^t \mathbf{A}_1 \mathbf{d} + \mathbf{p}_2^t \mathbf{A}_2 \mathbf{d} \ge \mathbf{p}_1^t \mathbf{z}_1 + \mathbf{p}_2^t \mathbf{z}_2 \qquad \text{for each } \mathbf{d} \in R^n \text{ and } (\mathbf{z}_1, \mathbf{z}_2) \in \text{cl}\, S_2.$$

Letting $\mathbf{z}_2 = \mathbf{0}$ and since each component of $\mathbf{z}_1$ can be made an arbitrarily large negative number, it follows that $\mathbf{p}_1 \ge \mathbf{0}$. Also, letting $(\mathbf{z}_1, \mathbf{z}_2) = (\mathbf{0}, \mathbf{0})$, we must have $(\mathbf{p}_1^t \mathbf{A}_1 + \mathbf{p}_2^t \mathbf{A}_2)\mathbf{d} \ge 0$ for each $\mathbf{d} \in R^n$. Letting $\mathbf{d} = -(\mathbf{A}_1^t \mathbf{p}_1 + \mathbf{A}_2^t \mathbf{p}_2)$, it follows that $-\left\| \mathbf{A}_1^t \mathbf{p}_1 + \mathbf{A}_2^t \mathbf{p}_2 \right\|^2 \ge 0$, and thus $\mathbf{A}_1^t \mathbf{p}_1 + \mathbf{A}_2^t \mathbf{p}_2 = \mathbf{0}$.

To summarize, we have shown that there exists a nonzero vector $\mathbf{p}^t = (\mathbf{p}_1^t, \mathbf{p}_2^t)$ with $\mathbf{p}_1 \ge \mathbf{0}$ such that $\mathbf{A}_1^t \mathbf{p}_1 + \mathbf{A}_2^t \mathbf{p}_2 = \mathbf{0}$. Denoting the components of $\mathbf{p}_1$ by u_0 and u_i for $i \in I$, and letting $\mathbf{p}_2 = \mathbf{v}$, the first result follows. The equivalent form of the necessary conditions is readily obtained by letting $u_i = 0$ for $i \notin I$, and the proof is complete.

The reader may note that the Lagrangian multiplier v_i associated with the ith equality constraints is unrestricted in sign. Note also that these conditions are *not* equivalently obtained by writing each equality as two associated inequalities and then applying the FJ conditions for the inequality-constrained case. The FJ conditions can also be written in vector notation as follows:

$$\begin{aligned} u_0 \nabla f(\overline{\mathbf{x}}) + \nabla \mathbf{g}(\overline{\mathbf{x}})^t \mathbf{u} + \nabla \mathbf{h}(\overline{\mathbf{x}})^t \mathbf{v} &= \mathbf{0} \\ \mathbf{u}^t \mathbf{g}(\overline{\mathbf{x}}) &= 0 \\ (u_0, \mathbf{u}) &\ge (0, \mathbf{0}) \\ (u_0, \mathbf{u}, \mathbf{v}) &\ne (0, \mathbf{0}, \mathbf{0}). \end{aligned}$$

Here $\nabla \mathbf{g}(\overline{\mathbf{x}})$ is an $m \times n$ Jacobian matrix whose ith row is $\nabla g_i(\overline{\mathbf{x}})^t$, and $\nabla \mathbf{h}(\overline{\mathbf{x}})$ is an $\ell \times n$ Jacobian matrix whose ith row is $\nabla h_i(\overline{\mathbf{x}})^t$. Also, $\mathbf{u}$ and $\mathbf{v}$ are, respectively, an m-vector and an ℓ-vector, denoting the Lagrangian multipliers associated with the inequality and equality constraints.

4.3.3 Example

$$\begin{aligned} \text{Minimize } & x_1^2 + x_2^2 \\ \text{subject to } & x_1^2 + x_2^2 \le 5 \\ & -x_1 \le 0 \\ & -x_2 \le 0 \\ & x_1 + 2x_2 = 4. \end{aligned}$$

Here, we have only one equality constraint. We verify below that the Fritz John conditions hold true at the optimal point $\bar{\mathbf{x}} = (4/5, 8/5)^t$. First, note that there are no binding inequality constraints at $\bar{\mathbf{x}}$; that is, $I = \varnothing$. Hence, the multipliers associated with the inequality constraints are equal to zero. Note that

$$\nabla f(\bar{\mathbf{x}}) = (8/5, 16/5)^t \qquad \text{and} \qquad \nabla h_1(\bar{\mathbf{x}}) = (1,2)^t.$$

Thus,

$$u_0 \begin{pmatrix} \frac{8}{5} \\ \frac{16}{5} \end{pmatrix} + v_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

is satisfied, for example, by $u_0 = 5$ and $v_1 = -8$.

4.3.4 Example

$$\begin{aligned} \text{Minimize } & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{subject to } & x_1^2 + x_2^2 \le 5 \\ & -x_1 \le 0 \\ & -x_2 \le 0 \\ & x_1 + 2x_2 = 4. \end{aligned}$$

This example is the same as Example 4.2.9, with the inequality constraint $x_1 + 2x_2 \le 4$ replaced by $x_1 + 2x_2 = 4$. At the optimal point $\bar{\mathbf{x}} = (2,1)^t$, we have only one inequality constraint $x_1^2 + x_2^2 \le 5$ binding. The Fritz John condition

$$u_0 \begin{pmatrix} -2 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + v_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

is satisfied, for example, by $u_0 = 3$, $u_1 = 1$, and $v_1 = 2$.

4.3.5 Example

$$\begin{aligned} \text{Minimize } & -x_1 \\ \text{subject to } & x_2 - (1 - x_1)^3 = 0 \\ & -x_2 - (1 - x_1)^3 = 0. \end{aligned}$$

As shown in Figure 4.12, this problem has only one feasible point, namely, $\bar{\mathbf{x}} = (1,0)^t$. At this point, we have

$$\nabla f(\bar{\mathbf{x}}) = (-1,0)^t, \qquad \nabla h_1(\bar{\mathbf{x}}) = (0,1)^t, \qquad \nabla h_2(\bar{\mathbf{x}}) = (0,-1)^t.$$

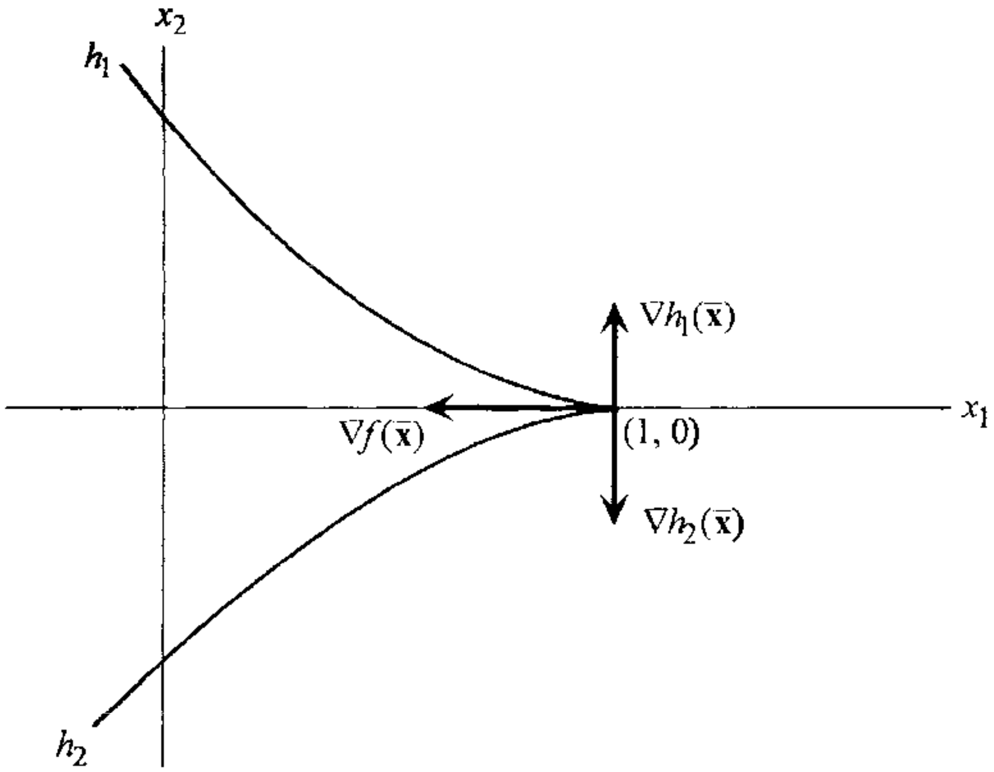


Figure 4.12 Setup for Example 4.3.5.

The condition

$$u_0\begin{pmatrix}-1\\0\end{pmatrix}+v_1\begin{pmatrix}0\\1\end{pmatrix}+v_2\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

holds true only if $u_0 = 0$ and $v_1 = v_2 = \alpha$, where α is any scalar. Thus, the Fritz John necessary conditions are met at the point $\bar{\mathbf{x}}$.

Similar to Theorem 4.2.12, we now provide a set of sufficient conditions that enable us to guarantee that a FJ point is a local minimum. Again, as with Theorem 4.2.12, several variations of such sufficient conditions are possible. We state the result below using one such condition motivated by the converse to Theorem 4.3.1, and ask the reader to explore other conditions in Exercise 4.22.

4.3.6 Theorem (Fritz John Sufficient Conditions)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$, g_i: $R^n \to R$, $i = 1,\ldots, m$, and h_i: $R^n \to R$, $i = 1,\ldots, \ell$. Consider the Problem P given below:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. & \end{array}$$

Let $\bar{\mathbf{x}}$ be a FJ solution and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Define $S = \{\mathbf{x} : g_i(\bar{\mathbf{x}}) \le 0$ for $i \in I$, $h_i(\bar{\mathbf{x}}) = 0$, for $i = 1, ..., \ell\}$. If h_i for $i = 1,..., \ell$ are affine and $\nabla h_i(\bar{\mathbf{x}})$, $i = 1,..., \ell$, are linearly independent, and if there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ of $\bar{\mathbf{x}}$, $\varepsilon > 0$ such that f is pseudoconvex on $S \cap N_\varepsilon(\bar{\mathbf{x}})$, and g_i for $i \in I$ are strictly pseudoconvex over $S \cap N_\varepsilon(\bar{\mathbf{x}})$, $\bar{\mathbf{x}}$ is a local minimum for Problem P.

Proof

Let us first show that $F_0 \cap G_0 \cap H_0 = \varnothing$, where these sets are as defined in Theorem 4.3.1. On the contrary, suppose that there exists a solution $\mathbf{d} \in F_0 \cap G_0 \cap H_0$. Then, by taking the inner product of the dual feasibility condition $u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$ with $\mathbf{d}$, we obtain $u_0 \nabla f(\bar{\mathbf{x}})^t \mathbf{d} + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} = 0$, since $\mathbf{d} \in H_0$. But $\mathbf{d} \in F_0 \cap G_0$ and $(u_0, u_i$ for $i \in I) \ge \mathbf{0}$ implies that $(u_0, u_i$ for $i \in I) = (0, \mathbf{0})$. Since $\bar{\mathbf{x}}$ is a FJ point, we therefore must have a solution to the system $\sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$, $\mathbf{v} \ne \mathbf{0}$, which contradicts the linear independence of $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,..., \ell$. Hence, $F_0 \cap G_0 \cap H_0 = \varnothing$.

Now, closely following the proof to the converse statement of Theorem 4.3.1, and restricting attention to $S \cap N_\varepsilon(\bar{\mathbf{x}})$, we can conclude that $\bar{\mathbf{x}}$ is a local minimum for P. This completes the proof.

Karush–Kuhn–Tucker Conditions

In the Fritz John conditions, the Lagrangian multiplier associated with the objective function is not necessarily positive. Under further assumptions on the constraint set, we can claim that at any local minimum, there exists a set of Lagrange multipliers for which u_0 is positive. In Theorem 4.3.7, we obtain a generalization of the KKT necessary optimality conditions of Theorem 4.2.13. This is done by imposing a qualification on the gradients of the equality and binding inequality constraints that ensure that $u_0 > 0$ necessarily holds true in the Fritz John conditions. Other qualifications on the constraints to ensure the existence of $u_0 > 0$ in the FJ conditions at a local minimum are discussed in Chapter 5.

4.3.7 Theorem (Karush–Kuhn–Tucker Necessary Conditions)

Let X be a nonempty open set in R^n, and let $f: R^n \to R$, $g_i: R^n \to R$ for $i = 1,..., m$, and $h_i: R^n \to R$ for $i = 1,..., \ell$. Consider the Problem P given below:

$$
\begin{aligned}
\text{Minimize} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_i(\mathbf{x}) \le 0 && \text{for } i = 1,\dots,m \\
& h_i(\mathbf{x}) = 0 && \text{for } i = 1,\dots,\ell \\
& \mathbf{x} \in X.
\end{aligned}
$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$, that each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and that each h_i for $i = 1,\dots,\ell$ is continuously differentiable at $\bar{\mathbf{x}}$. Further, suppose that $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\dots,\ell$ are linearly independent. (Such an $\bar{\mathbf{x}}$ is sometimes called *regular*.) If $\bar{\mathbf{x}}$ solves Problem P locally, there exist unique scalars u_i for $i \in I$ and v_i for $i = 1,\dots,\ell$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i \ge 0 \qquad \text{for } i \in I.$$

In addition to the above assumptions, if each g_i for $i \notin I$ is also differentiable at $\bar{\mathbf{x}}$, the KKT conditions can be written in the following equivalent form:

$$
\begin{aligned}
\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) &= \mathbf{0} \\
u_i g_i(\bar{\mathbf{x}}) &= 0 && \text{for } i = 1,\dots,m \\
u_i &\ge 0 && \text{for } i = 1,\dots,m.
\end{aligned}
$$

Proof

By Theorem 4.3.2 there exist scalars u_0 and $\hat{u}_i$ for $i \in I$, and $\hat{v}_i$ for $i = 1,\dots,\ell$, not all zero, such that

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} \hat{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \hat{v}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0} \tag{4.15}$$

$$u_0, \hat{u}_i \ge 0 \qquad \text{for } i \in I.$$

Note that $u_0 > 0$, because if $u_0 = 0$, (4.15) would contradict the assumption of linear independence of $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\dots,\ell$. The first result then follows by letting $u_i = \hat{u}_i/u_0$ for $i \in I$, and $v_i = \hat{v}_i/u_0$ for $i = 1,\dots,\ell$, and noting that the linear independence assumption implies the uniqueness of these Lagrangian multipliers. The equivalent form of the necessary conditions follows by letting $u_i = 0$ for $i \notin I$. This completes the proof.

Note that the KKT conditions of Theorem 4.3.7 can be written in vector form as follows:

$$\nabla f(\bar{\mathbf{x}}) + \nabla \mathbf{g}(\bar{\mathbf{x}})^t \mathbf{u} + \nabla \mathbf{h}(\bar{\mathbf{x}})^t \mathbf{v} = \mathbf{0}$$
$$\mathbf{u}^t \mathbf{g}(\bar{\mathbf{x}}) = 0$$
$$\bar{\mathbf{u}} \geq \mathbf{0}.$$

Here $\nabla \mathbf{g}(\bar{\mathbf{x}})$ is an $m \times n$ Jacobian matrix and $\nabla \mathbf{h}(\bar{\mathbf{x}})$ is an $\ell \times n$ Jacobian matrix whose ith rows, respectively, are $\nabla g_i(\bar{\mathbf{x}})^t$ and $\nabla h_i(\bar{\mathbf{x}})^t$. The vectors u and v are the Lagrangian multiplier vectors.

The reader might have observed that the KKT conditions of Theorem 4.3.6 are precisely the KKT conditions of the inequality case given in Theorem 4.2.13 when each equality constraint $h_i(\mathbf{x}) = 0$ is replaced by the two equivalent inequalities $h_i(\mathbf{x}) \leq 0$ and $-h_i(\mathbf{x}) \leq 0$, for $i = 1,\ldots, \ell$. Denoting v_i^+ and v_i^- as the nonnegative Lagrangian multipliers associated with the latter two inequalities and using the KKT conditions for the inequality case produces the KKT conditions of Theorem 4.3.7 upon replacing the difference $v_i^+ - v_i^-$ of two nonnegative variables by the unrestricted variable v_i for each $i = 1,\ldots, \ell$. In fact, writing the equalities as equivalent inequalities, the sets G_0' and G' defined in Theorem 4.2.15 become, respectively, $G_0' \cap H_0$ and $G' \cap H_0$. Theorem 4.2.15 then asserts that for Problem P of the present section,

$$\bar{\mathbf{x}} \text{ is a KKT solution} \Leftrightarrow F_0 \cap G_0' \cap H_0 = \varnothing \Leftrightarrow F_0 \cap G' \cap H_0 = \varnothing. \quad (4.16)$$

Moreover, this happens if and only if $\bar{\mathbf{x}}$ solves the first-order linear programming approximation LP($\bar{\mathbf{x}}$) at the point $\bar{\mathbf{x}}$ given by

$$\text{LP}(\bar{\mathbf{x}}): \text{ Minimize}\{f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) : g_i(\bar{\mathbf{x}}) + \nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \leq 0$$
$$\text{for } i = 1,\ldots, m,\ \nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) = 0 \text{ for } i = 1,\ldots, \ell\}. \quad (4.17)$$

Now consider Examples 4.3.3, 4.3.4, and 4.3.5. In Example 4.3.3 the reader can verify that $u_1 = u_2 = u_3 = 0$ and $v_1 = -8/5$ satisfy the KKT conditions at $\bar{\mathbf{x}} = (4/5, 8/5)^t$. In Example 4.3.4, the values of the multipliers satisfying the KKT conditions at $\bar{\mathbf{x}} = (2,1)^t$ are

$$u_1 = 1/3, \qquad u_2 = u_3 = 0, \qquad v_1 = 2/3.$$

Finally, Example 4.3.5 does not satisfy the constraint qualification of Theorem 4.3.7 at $\bar{\mathbf{x}} = (1,0)^t$, since $\nabla h_1(\bar{\mathbf{x}})$ and $\nabla h_2(\bar{\mathbf{x}})$ are linearly dependent. In fact, no constraint qualification (known or unknown!) can hold true at this point $\bar{\mathbf{x}}$ because it is not a KKT point. The feasible region for the first-order linear programming approximation LP($\bar{\mathbf{x}}$) is given by the entire x_1-axis; and unless $\nabla f(\bar{\mathbf{x}})$ is orthogonal to this axis, $\bar{\mathbf{x}}$ is not an optimal solution for LP($\bar{\mathbf{x}}$).

Theorem 4.3.8 shows that under rather mild convexity assumptions on f, g_i, and h_i, the KKT conditions are also sufficient for local optimality. Again, we fashion this result following Theorem 4.2.16 and ask the reader to investigate other variations in Exercises 4.22 and 4.50.

4.3.8 Theorem (Karush–Kuhn–Tucker Sufficient Conditions)

Let X be a nonempty open set in R^n, and let f: $R^n \to R$, g_i: $R^n \to R$ for $i = 1,\ldots,m$, and h_i: $R^n \to R$ for $i = 1,\ldots,\ell$. Consider Problem P:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) & \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 & \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. & \end{aligned}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that the KKT conditions hold true at $\bar{\mathbf{x}}$; that is, there exist scalars $\bar{u}_i \ge 0$ for $i \in I$ and $\bar{v}_i$ for $i = 1,\ldots,\ell$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}. \tag{4.18}$$

Let $J = \{i : \bar{v}_i > 0\}$ and $K = \{i : \bar{v}_i < 0\}$. Further, suppose that f is pseudoconvex at $\bar{\mathbf{x}}$, g_i is quasiconvex at $\bar{\mathbf{x}}$ for $i \in I$, h_i is quasiconvex at $\bar{\mathbf{x}}$ for $i \in J$, and h_i is quasiconcave at $\bar{\mathbf{x}}$ for $i \in K$. Then $\bar{\mathbf{x}}$ is a global optimal solution to Problem P. In particular, if the generalized convexity assumptions on the objective and constraint functions are restricted to the domain $N_\varepsilon(\bar{\mathbf{x}})$ for some $\varepsilon > 0$, $\bar{\mathbf{x}}$ is a local minimum for P.

Proof

Let $\mathbf{x}$ be any feasible solution to Problem P. [In case the generalized convexity assumptions hold true only by restricting the domain of the objective and constraint functions to $N_\varepsilon(\bar{\mathbf{x}})$, let $\mathbf{x}$ be any feasible solution to Problem P that also lies within $N_\varepsilon(\bar{\mathbf{x}})$.]

Then, for $i \in I$, $g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}})$, since $g_i(\mathbf{x}) \le 0$ and $g_i(\bar{\mathbf{x}}) = 0$. By the quasiconvexity of g_i at $\bar{\mathbf{x}}$ it follows that

$$g_i(\bar{\mathbf{x}} + \lambda(\mathbf{x} - \bar{\mathbf{x}})) = g_i(\lambda \mathbf{x} + (1-\lambda)\bar{\mathbf{x}}) \le \max\{g_i(\mathbf{x}), g_i(\bar{\mathbf{x}})\} = g_i(\bar{\mathbf{x}})$$

for all $\lambda \in (0, 1)$. This implies that g_i does not increase by moving from $\bar{\mathbf{x}}$ along the direction $\mathbf{x} - \bar{\mathbf{x}}$. Thus, by Theorem 4.1.2 we must have

$$\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \le 0 \qquad \text{for } i \in I. \tag{4.19}$$

Similarly, since h_i is quasiconvex at $\bar{\mathbf{x}}$ for $i \in J$, and h_i is quasiconcave at $\bar{\mathbf{x}}$ for $i \in K$, we have

$$\nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \le 0 \qquad \text{for } i \in J \tag{4.20}$$

$$\nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \ge 0 \qquad \text{for } i \in K. \tag{4.21}$$

Multiplying (4.19), (4.20), and (4.21) by $\bar{u}_i \ge 0$, $\bar{v}_i > 0$, and $\bar{v}_i < 0$, respectively, and adding, we get

$$\left[\sum_{i \in I} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i \in J \cup K} \bar{v}_i \nabla h_i(\bar{\mathbf{x}})\right]^t (\mathbf{x} - \bar{\mathbf{x}}) \le 0. \tag{4.22}$$

Multiplying (4.18) by $\mathbf{x} - \bar{\mathbf{x}}$ and noting that $\bar{v}_i = 0$ for $i \notin J \cup K$, (4.22) implies that

$$\nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \ge 0.$$

By the pseudoconvexity of f at $\bar{\mathbf{x}}$, we must have $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$, the proof is complete.

It is instructive to note that as evident from Theorem 4.3.8 and its proof, the equality constraints having positive Lagrangian multipliers at $\bar{\mathbf{x}}$ can be replaced by "less than or equal to" constraints, and those having negative Lagrangian multipliers at $\bar{\mathbf{x}}$ can be replaced by "greater than or equal to" constraints, whereas those having zero Lagrangian multipliers can be deleted and $\bar{\mathbf{x}}$ will still remain a KKT solution for this relaxed problem P′, say. Hence, noting Theorem 4.2.16, the generalized convexity assumptions of Theorem 4.3.8 imply that $\bar{\mathbf{x}}$ is optimal to the relaxed problem P′, and being feasible to P, it is optimal for P (globally or locally as the case might be). This argument provides an alternative simpler proof for Theorem 4.3.8 based on Theorem 4.2.16. Moreover, it asserts that under generalized convexity assumptions, the sign of the Lagrangian multipliers can be used to assess whether an equality constraint is effectively behaving as a "less than or equal to" or a "greater than or equal to" constraint.

Two points of caution are worth noting here in connection with the foregoing relaxation P′ of P. First, under the (generalized) convexity assumptions, deleting an equality constraint that has a zero Lagrangian multiplier can create alternative optimal solutions that are not feasible to the original problem. For example, in the problem to minimize $\{x_1 : x_1 \ge 0 \text{ and } x_2 = 1\}$, the Lagrangian multiplier associated with the equality at the unique optimum $\bar{\mathbf{x}} = (0,1)^t$ is zero. However, deleting this constraint produces an infinite number of alternative optimal solutions.

Second, for the nonconvex case, note that even if $\bar{\mathbf{x}}$ is optimal for P, it may not even be a local optimum for P′, although it remains a KKT point for P′. For example, consider the problem to minimize $(-x_1^2 - x_2^2)$ subject to $x_1 = 0$ and $x_2 = 0$. The unique optimum is obviously $\bar{\mathbf{x}} = (0,0)^t$, and the Lagrangian multipliers associated with the constraints at $\bar{\mathbf{x}}$ are both zeros. However, deleting either of the two constraints, or even replacing either with a "less than or equal to" or "greater than or equal to" inequality, will make the problem unbounded. In general, the reader should bear in mind that deleting even nonbinding constraints for nonconvex problems can change the optimality status of a solution. Figure 4.13 illustrates one such situation. Here $g_2(\mathbf{x}) \le 0$ is nonbinding at the optimum $\bar{\mathbf{x}}$; but deleting it changes the global optimum to the point $\hat{\mathbf{x}}$, leaving $\bar{\mathbf{x}}$ as only a local minimum. (See Exercise 4.24 for an instance in which the optimum does not even remain locally optimum after deleting a nonbinding constraint.)

Alternative Forms of the Karush–Kuhn–Tucker Conditions for General Problems

Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$, $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$, and $\mathbf{x} \in X$, where X is an open set in R^n. We have derived above the following necessary conditions of optimality at a feasible point $\bar{\mathbf{x}}$ (under a suitable constraint qualification):

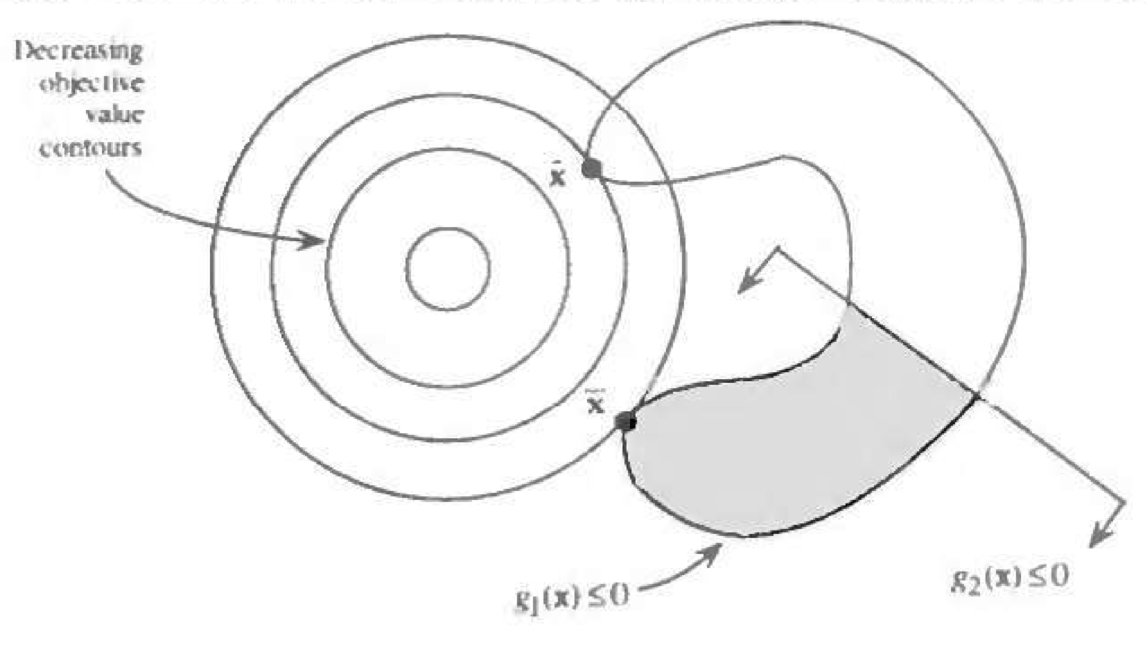


Figure 4.13 Caution on deleting nonbinding constraints for nonconvex problems.

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots,m.$$

Some authors prefer to use the multipliers $\lambda_i = -u_i \leq 0$ and $\mu_i = -v_i$. In this case, the KKT conditions can be written as follows:

$$\nabla f(\bar{\mathbf{x}}) - \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{\mathbf{x}}) - \sum_{i=1}^{\ell} \mu_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$\lambda_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$\lambda_i \leq 0 \qquad \text{for } i = 1,\ldots,m.$$

Now, consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \leq 0$ for $i = 1,\ldots,m_1$, $g_i(\mathbf{x}) \geq 0$ for $i = m_1+1,\ldots,m$, $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots,\ell$, and $\mathbf{x} \in X$, where X is an open set in R^n. Writing $g_i(\mathbf{x}) \geq 0$ for $i = m_1+1,\ldots,m$ as $-g_i(\mathbf{x}) \leq 0$ for $i = m_1+1,\ldots,m$, and using the results of Theorem 4.3.7, the necessary conditions for this problem can be expressed as follows:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots,m_1$$

$$u_i \leq 0 \qquad \text{for } i = m_1+1,\ldots,m.$$

We now consider problems of the type to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \leq 0$ for $i = 1,\ldots, m$, $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots,\ell$, and $\mathbf{x} \geq \mathbf{0}$. Such problems with nonnegativity restrictions on the variables frequently arise in practice. Clearly, the KKT conditions discussed earlier would apply as usual. However, it is sometimes convenient to eliminate the Lagrangian multipliers associated with $\mathbf{x} \geq \mathbf{0}$. The conditions then reduce to

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) \geq \mathbf{0}$$

$$\left[\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}})\right]^t \bar{\mathbf{x}} = 0$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots,m.$$

Finally, consider the problem to maximize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots,m_1$, $g_i(\mathbf{x}) \ge 0$ for $i = m_1+1,\ldots,m$, $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$, and $\mathbf{x} \in X$, where X is an open set in R^n. The necessary conditions for optimality can be written as follows:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$u_i \le 0 \qquad \text{for } i = 1,\ldots,m_1$$

$$u_i \ge 0 \qquad \text{for } i = m_1+1,\ldots,m.$$

4.4 Second-Order Necessary and Sufficient Conditions for Constrained Problems

In Section 4.1 we considered the unconstrained problem of minimizing $f(\mathbf{x})$ subject to $\mathbf{x} \in R^n$, and assuming differentiability, we derived the first-order necessary optimality condition that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ at all local optimal solutions $\bar{\mathbf{x}}$. However, when $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, $\bar{\mathbf{x}}$ can be a local minimum, a local maximum, or an inflection point. To further reduce the candidate set of solutions produced by this first-order necessary optimality condition, and to assess the local optimality status of a given candidate solution, we developed second-order (and higher) necessary and/or sufficient optimality conditions.

Over Sections 4.2 and 4.3 we have developed first-order necessary optimality conditions for constrained problems. In particular, assuming a suitable constraint qualification, we have derived the first-order necessary KKT optimality conditions. Based on various (generalized) convexity assumptions, we have provided sufficient conditions to guarantee that a given solution that satisfies the first-order optimality conditions is globally or locally optimum. Analogous to the unconstrained case, we now derive second-order necessary and sufficient optimality conditions for constrained problems.

Toward this end, let us introduce the concept of a *Lagrangian function.* Consider the problem:

$$\text{P: Minimize}\{f(\mathbf{x}) : \mathbf{x} \in S\}, \tag{4.23a}$$

where

$$S = \{\mathbf{x} : g_i(\mathbf{x}) \le 0 \text{ for } i = 1,\ldots,m,\ h_i(\mathbf{x}) = 0 \text{ for } i = 1,\ldots,\ell, \text{ and } \mathbf{x} \in X\}. \tag{4.23b}$$

Assume that f, g_i for $i = 1,\ldots, m$, and h_i for $i = 1,\ldots, \ell$ are all defined on $R^n \to R$ and are twice differentiable, and that X is a nonempty open set in R^n. The *Lagrangian function* for this problem is defined as

$$\phi(\mathbf{x},\mathbf{u},\mathbf{v}) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} v_i h_i(\mathbf{x}). \tag{4.24}$$

As we shall learn in Chapter 6, this function enables us to formulate a duality theory for nonlinear programming problems, akin to that for linear programming problems as expounded in Theorem 2.7.3 and its corollaries. Now, let $\bar{\mathbf{x}}$ be a KKT point for Problem P, with associated Lagrangian multipliers $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ corresponding to the inequality and equality constraints, respectively. Conditioned on $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$, define the *restricted Lagrangian function*

$$L(\mathbf{x}) \equiv \phi(\mathbf{x},\bar{\mathbf{u}},\bar{\mathbf{v}}) = f(\mathbf{x}) + \sum_{i\in I} \bar{u}_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} \bar{v}_i h_i(\mathbf{x}), \tag{4.25}$$

where $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$ is the index set of the binding inequality constraints at $\bar{\mathbf{x}}$.

Observe that the dual feasibility condition

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i\in I} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0} \tag{4.26}$$

in the KKT system is equivalent to the statement that the gradient, $\nabla L(\bar{\mathbf{x}})$, of L at $\mathbf{x} = \bar{\mathbf{x}}$ vanishes. Moreover, we have

$$L(\mathbf{x}) \le f(\mathbf{x}) \qquad \text{for all } \mathbf{x} \in S, \qquad \text{while } L(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) \tag{4.27}$$

because $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots,\ell$ and $g_i(\mathbf{x}) \le 0$ for $i \in I$ for all $\mathbf{x} \in S$, while $\bar{u}_i g_i(\bar{\mathbf{x}}) = 0$ for $i \in I$, and $h_i(\bar{\mathbf{x}}) = 0$ for $i = 1,\ldots,\ell$. Hence, if $\bar{\mathbf{x}}$ turns out to be a (local) minimizer for L, it will also be a (local) minimizer for Problem P. This is formalized below.

4.4.1 Lemma

Consider Problem P as defined in (4.23), where the objective and constraint defining functions are all twice differentiable, and where X is a nonempty, open set in R^n. Suppose that $\bar{\mathbf{x}}$ is a KKT point for Problem P with Lagrangian multipliers $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ associated with the inequality and equality constraints, respectively. Define the restricted Lagrangian function L as in (4.25), and denote its Hessian by $\nabla^2 L$.

a. If $\nabla^2 L$ is positive semidefinite for all $\mathbf{x} \in S$, $\bar{\mathbf{x}}$ is a global minimum for Problem P. On the other hand, if $\nabla^2 L$ is positive semidefinite for all $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$ for some ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ about $\bar{\mathbf{x}}$, $\varepsilon > 0$, $\bar{\mathbf{x}}$ is a local minimum for Problem P.

b. If $\nabla^2 L(\bar{\mathbf{x}})$ is positive definite, $\bar{\mathbf{x}}$ is a strict local minimum for Problem P.

Proof

From (4.25) and (4.26), we have that $\nabla L(\bar{\mathbf{x}}) = \mathbf{0}$. Hence, under the first condition of Part a, we obtain, by the convexity of $L(\mathbf{x})$ over S, that $L(\bar{\mathbf{x}}) \le L(\mathbf{x})$ for all $\mathbf{x} \in S$; and thus from (4.27) we get $f(\bar{\mathbf{x}}) = L(\bar{\mathbf{x}}) \le L(\mathbf{x}) \le f(\mathbf{x})$ for all $\mathbf{x} \in S$. Therefore, $\bar{\mathbf{x}}$ solves Problem P. By restricting attention to $S \cap N_\varepsilon(\bar{\mathbf{x}})$ in the second case of Part a, we conclude similarly that $f(\bar{\mathbf{x}}) \le f(\mathbf{x})$ for all $\mathbf{x} \in S \cap N_\varepsilon(\bar{\mathbf{x}})$. This proves Part a.

Similarly, if $\nabla^2 L(\bar{\mathbf{x}})$ is positive definite, by Theorem 4.1.4 since $\nabla L(\bar{\mathbf{x}}) = \mathbf{0}$, we have that $\bar{\mathbf{x}}$ is a strict local minimum for L. Hence, from (4.27) we deduce that $f(\bar{\mathbf{x}}) = L(\bar{\mathbf{x}}) < L(\mathbf{x}) \le f(\mathbf{x})$ for all $\mathbf{x} \ne \bar{\mathbf{x}}$ in $S \cap N_\varepsilon(\bar{\mathbf{x}})$ for some ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ of $\bar{\mathbf{x}}$, $\varepsilon > 0$, and this completes the proof.

The above result is related to the *saddle point optimality conditions* explored more fully in Chapter 6, which establish that a KKT solution $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for which $\bar{\mathbf{x}}$ minimizes $L(\mathbf{x})$ subject to $\mathbf{x} \in S$ corresponds to a certain pair of primal and dual problems having no duality gap. Indeed, observe that the global (or local) optimality claims in Lemma 4.4.1 continue to hold true under the less restrictive assumption that $\bar{\mathbf{x}}$ globally (or locally) minimizes $L(\mathbf{x})$ over S. However, our choice of stating Lemma 4.4.1 as above is motivated by the following result, which asserts that $\mathbf{d}^t \nabla^2 L(\bar{\mathbf{x}}) \mathbf{d}$ needs to be positive only for $\mathbf{d}$ restricted to lie in a specified cone rather than for all $\mathbf{d} \in R^n$ as in Lemma 4.4.1b, for us to be able to claim that $\bar{\mathbf{x}}$ is a strict local minimum for P. In other words, this result is shown to hold true whenever the Lagrangian function $L(\mathbf{x})$ displays a positive curvature at $\bar{\mathbf{x}}$ along directions restricted to the set given below.

4.4.2 Theorem (KKT Second-Order Sufficient Conditions)

Consider Problem P as defined in (4.23), where the objective and constraint defining functions are all twice differentiable, and where X is a nonempty, open set in R^n. Let $\bar{\mathbf{x}}$ be a KKT point for Problem P, with Lagrangian multipliers $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ associated with the inequality and equality constraints, respectively. Let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$, and denote $I^+ = \{i \in I : \bar{u}_i > 0\}$ and $I^0 = \{i \in I : \bar{u}_i = 0\}$. ($I^+$ and I^0 are sometimes referred to as the set of *strongly active* and *weakly active constraints*, respectively.) Define the restricted Lagrangian function $L(\mathbf{x})$ as in (4.25), and denote its Hessian at $\bar{\mathbf{x}}$ by

$$\nabla^2 L(\bar{\mathbf{x}}) \equiv \nabla^2 f(\bar{\mathbf{x}}) + \sum_{i \in I} \bar{u}_i \nabla^2 g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla^2 h_i(\bar{\mathbf{x}}),$$

where $\nabla^2 f(\bar{\mathbf{x}})$, $\nabla^2 g_i(\bar{\mathbf{x}})$ for $i \in I$, and $\nabla^2 h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots,\ell$, are the Hessians of f, g_i for $i \in I$, and h_i for $i = 1,\ldots,\ell$, respectively, all evaluated at $\bar{\mathbf{x}}$. Define the cone

$$\begin{aligned} C = \{\mathbf{d} \neq \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \quad & \text{for } i \in I^+, \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \leq 0 \text{ for } i \in I^0, \\ \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \quad & \text{for } i = 1,\ldots,\ell\}. \end{aligned}$$

Then if $\mathbf{d}^t \nabla^2 L(\bar{\mathbf{x}})\mathbf{d} > 0$ for all $\mathbf{d} \in C$, we have that $\bar{\mathbf{x}}$ is a strict local minimum for P.

Proof

Suppose that $\bar{\mathbf{x}}$ is not a strict local minimum. Then, as in Theorem 4.1.4, there exists a sequence $\{\mathbf{x}_k\}$ in S converging to $\bar{\mathbf{x}}$ such that $\mathbf{x}_k \neq \bar{\mathbf{x}}$ and $f(\mathbf{x}_k) \leq f(\bar{\mathbf{x}})$ for all k. Defining $\mathbf{d}_k = (\mathbf{x}_k - \bar{\mathbf{x}})/\|\mathbf{x}_k - \bar{\mathbf{x}}\|$ and $\lambda_k = \|\mathbf{x}_k - \bar{\mathbf{x}}\|$ for all k, we have that $\mathbf{x}_k = \bar{\mathbf{x}} + \lambda_k \mathbf{d}_k$, where $\|\mathbf{d}_k\| = 1$ for all k, and we have that $\{\lambda_k\} \to 0^+$ as $k \to \infty$. Since $\|\mathbf{d}_k\| = 1$ for all k, a convergent subsequence exists. Assume, without loss of generality, that the given sequence itself represents this convergent subsequence. Hence, $\{\mathbf{d}_k\} \to \mathbf{d}$, where $\|\mathbf{d}_k\| = 1$. Moreover, we have

$$\begin{aligned} 0 \geq\ & f(\bar{\mathbf{x}} + \lambda_k \mathbf{d}_k) - f(\bar{\mathbf{x}}) = \lambda_k \nabla f(\bar{\mathbf{x}})^t \mathbf{d}_k \\ +\ & (1/2)\lambda_k^2 \mathbf{d}_k^t \nabla^2 f(\bar{\mathbf{x}})\mathbf{d}_k + \lambda_k^2 \alpha_f(\bar{\mathbf{x}};\lambda_k \mathbf{d}_k), \end{aligned} \tag{4.28a}$$

$$\begin{aligned} 0 \geq\ & g_i(\bar{\mathbf{x}} + \lambda_k \mathbf{d}_k) - g_i(\bar{\mathbf{x}}) = \lambda_k \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d}_k \\ +\ & (1/2)\lambda_k^2 \mathbf{d}_k^t \nabla^2 g_i(\bar{\mathbf{x}})\mathbf{d}_k + \lambda_k^2 \alpha_{g_i}(\bar{\mathbf{x}};\lambda_k \mathbf{d}_k) \quad \text{for } i \in I, \end{aligned} \tag{4.28b}$$

$$\begin{aligned} 0 =\ & h_i(\bar{\mathbf{x}} + \lambda_k \mathbf{d}_k) - h_i(\bar{\mathbf{x}}) = \lambda_k \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d}_k \\ +\ & (1/2)\lambda_k^2 \mathbf{d}_k^t \nabla^2 h_i(\bar{\mathbf{x}})\mathbf{d}_k + \lambda_k^2 \alpha_{h_i}(\bar{\mathbf{x}};\lambda_k \mathbf{d}_k) \quad \text{for } i \in 1,\ldots,\ell, \end{aligned} \tag{4.28c}$$

where α_f, α_{g_i} for $i \in I$, and α_{h_i} for $i = 1,\ldots,\ell$, all approach zero as $k \to \infty$. Dividing each expression in (4.28) by $\lambda_k > 0$ and taking limits as $k \to \infty$, we obtain

$$\nabla f(\bar{\mathbf{x}})^t \mathbf{d} \leq 0, \;\; \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \leq 0 \text{ for } i \in I \quad \text{and}$$

$$\nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\ldots,\ell. \tag{4.29}$$

Now, since $\bar{\mathbf{x}}$ is a KKT point, we have $\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$. Taking the inner product of this with $\mathbf{d}$ and using (4.29), we conclude that

$$\nabla f(\bar{\mathbf{x}})^t\mathbf{d} = 0, \qquad \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} = 0 \text{ for } i \in I^+, \qquad \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0 \text{ for } i \in I^0,$$

$$\text{and} \qquad \nabla h_i(\bar{\mathbf{x}})^t\mathbf{d} = 0 \text{ for } i = 1,\ldots,\ell. \tag{4.30}$$

Hence, in particular, d $\in C$. Furthermore, multiplying each of (4.28b) by $\bar{u}_i$ for $i \in I$, and each of (4.28c) by $\bar{v}_i$ for $i = 1,\ldots,\ell$, and adding, we get, using $\nabla f(\bar{\mathbf{x}})^t\mathbf{d}_k + \sum_{i\in I}\bar{u}_i\nabla g_i(\bar{\mathbf{x}})^t\mathrm{d}_k + \sum_{i=1}^{\ell}\bar{v}_i\nabla h_i(\bar{\mathbf{x}})^t\mathbf{d}_k = 0$,

$$0 \ge \frac{\lambda_k^2}{2}\mathbf{d}_k^t\nabla^2 L(\bar{\mathbf{x}})\mathbf{d}_k$$

$$+\lambda_k^2\left[\alpha_f(\bar{\mathbf{x}};\lambda_k\mathbf{d}_k) + \sum_{i\in I}\bar{u}_i\alpha_{g_i}(\bar{\mathbf{x}};\lambda_k\mathbf{d}_k) + \sum_{i=1}^{\ell}\bar{v}_i\alpha_{h_i}(\bar{\mathbf{x}};\lambda_k\mathbf{d}_k)\right].$$

Dividing the above inequality by $\lambda_k^2 > 0$ and taking limits as $k \to \infty$, we obtain $\mathbf{d}^t\nabla^2 L(\bar{\mathbf{x}})\mathbf{d} \le 0$, where $\|\mathbf{d}\| = 1$ and $\mathbf{d} \in C$. This is a contradiction. Therefore, $\bar{\mathbf{x}}$ must be a strict local minimum for Problem P, and the proof is complete.

Corollary

Consider Problem P as defined in the theorem, and let $\bar{\mathbf{x}}$ be a KKT point with associated Lagrangian multipliers $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ corresponding to the inequality and equality constraints, respectively. Furthermore, suppose that the collection of vectors $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I^+ = \{i \in I : \bar{u}_i > 0\}$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots,\ell$ contains a set of n linearly independent vectors. Then $\bar{\mathbf{x}}$ is a strict local minimum for P.

Proof

Under the stated linear independence condition of the corollary, we have that $C = \varnothing$, so Theorem 4.4.2 holds true vacuously by default. This completes the proof.

Several remarks concerning Theorem 4.4.2 are in order at this point. First, observe that it might appear from the proof of the theorem that the result can be strengthened by further restricting the cone C to include the constraint $\nabla f(\bar{\mathbf{x}})^t\mathrm{d} = 0$. Although this is valid, it does not further restrict C, since when $\bar{\mathbf{x}}$ is a KKT point and d $\in C$, we have automatically that $\nabla f(\bar{\mathbf{x}})^t\mathrm{d} = 0$. Second, observe that if the problem is unconstrained, Theorem 4.2.2 reduces to asserting that if $\nabla f(\bar{\mathbf{x}}) = 0$ and if $\nabla^2 f(\bar{\mathbf{x}}) \equiv \mathrm{H}(\bar{\mathbf{x}})$ is positive definite, $\bar{\mathbf{x}}$ is a strict local minimum. Hence, Theorem 4.1.4 is a special case of this result. Similarly, Lemma 4.4.1b is a special case of this result. Finally, observe that for linear programming problems, this sufficient condition does not necessarily hold true, except under the condition of the corollary, whence $\bar{\mathbf{x}}$ is a unique extreme point optimal solution.

We now turn our attention to the counterpart of Theorem 4.4.2 that deals with second-order necessary optimality conditions. Theorem 4.4.3 shows that if $\bar{\mathbf{x}}$ is a local minimum, then under a suitable second-order constraint qualification, it is a KKT point; and moreover, $\mathbf{d}^t\nabla^2 L(\bar{\mathbf{x}})\mathbf{d} \geq 0$ for all d belonging to C as defined in Theorem 4.4.2. The last statement indicates that the Lagrangian function L has a nonnegative curvature at $\bar{\mathbf{x}}$ along any direction in C.

4.4.3 Theorem (KKT Second-Order Necessary Conditions)

Consider Problem P as defined in (4.23), where the objective and constraint defining functions are all twice differentiable, and where X is a nonempty, open set in R^n. Let $\bar{\mathbf{x}}$ be a local minimum for Problem P, and denote $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Define the restricted Lagrangian function $L(\mathbf{x})$ as in (4.25), and denote its Hessian at $\bar{\mathbf{x}}$ by

$$\nabla^2 L(\bar{\mathbf{x}}) \equiv \nabla^2 f(\bar{\mathbf{x}}) + \sum_{i\in I} \bar{u}_i \nabla^2 g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla^2 h_i(\bar{\mathbf{x}}),$$

where $\nabla^2 f(\bar{\mathbf{x}})$, $\nabla^2 g_i(\bar{\mathbf{x}})$ for $i \in I$, and $\nabla^2 h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots,\ell$ are the Hessians of f, g_i for $i \in I$, and h_i for $i = 1,\ldots,\ell$, respectively, all evaluated at $\bar{\mathbf{x}}$. Assume that $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$, and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1\ldots,\ell$, are linearly independent. Then $\bar{\mathbf{x}}$ is a KKT point having Lagrange multipliers $\bar{\mathbf{u}} \geq \mathbf{0}$ and $\bar{\mathbf{v}}$ associated with the inequality and the equality constraints, respectively. Moreover, $\mathbf{d}^t\nabla^2 L(\bar{\mathbf{x}})\mathbf{d} \geq 0$ for all d $\in C = \{\mathbf{d} \neq \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$ for $i \in I^+$, $\nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \leq 0$ for $i \in I^0$, $\nabla h_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$ for all $i = 1,\ldots,\ell\}$, where $I^+ = \{i \in I : \bar{u}_i > 0\}$ and $I^0 = \{i \in I : \bar{u}_i = 0\}$.

Proof

By Theorem 4.3.7 we have directly that $\bar{\mathbf{x}}$ is a KKT point. Now, if $C = \emptyset$, the result is trivially true. Otherwise, consider any d $\in C$, and denote $I(\mathbf{d}) = \{i \in I : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} = 0\}$. For $\lambda \geq 0$, define $\alpha : R \to R^n$ by the following differential equation and boundary condition:

$$\frac{d\alpha(\lambda)}{d\lambda} = \mathrm{P}(\lambda)\mathbf{d}, \qquad \alpha(0) = \bar{\mathbf{x}},$$

where $\mathrm{P}(\lambda)$ is the matrix that projects any vector in the null space of the matrix having rows $\nabla g_i(\alpha(\lambda))$, $i \in I(\mathbf{d})$, and $\nabla h_i(\alpha(\lambda))$, $i = 1,\ldots,\ell$. Following the proof of Theorem 4.3.1 [by treating g_i for $i \in I(\mathbf{d})$, and h_i, $i = 1,\ldots,\ell$, as the "equations" therein, and treating g_i for $i \in I - I(\mathbf{d})$, for which $\nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} < 0$, as the "inequalities" therein], we obtain that $\boldsymbol{\alpha}(\lambda)$ is feasible for $0 \le \lambda \le \delta$, for some $\delta > 0$.

Now, consider a sequence $\{\lambda_k\} \to 0^+$ and denote $\mathbf{x}_k = \boldsymbol{\alpha}(\lambda_k)$ for all k. By the Taylor series expansion, we have

$$L(\mathbf{x}_k) = L(\overline{\mathbf{x}}) + \nabla L(\overline{\mathbf{x}})^t(\mathbf{x}_k - \overline{\mathbf{x}}) + \frac{1}{2}(\mathbf{x}_k - \overline{\mathbf{x}})^t \nabla^2 L(\overline{\mathbf{x}})(\mathbf{x}_k - \overline{\mathbf{x}})$$

$$+\|(\mathbf{x}_k - \overline{\mathbf{x}})\|^2 \beta[\overline{\mathbf{x}};(\mathbf{x}_k - \overline{\mathbf{x}})], \tag{4.31}$$

where $\beta[\overline{\mathbf{x}};(\mathbf{x}_k - \overline{\mathbf{x}})] \to 0$ as $\mathbf{x}_k \to \overline{\mathbf{x}}$. Since $g_i(\mathbf{x}_k) = 0$ for all $i \in I(\mathrm{d}) \supseteq I^+$ and $h_i(\mathbf{x}_k) = 0$ for all $i = 1,\ldots, \ell$, we have that $L(\mathbf{x}_k) = f(\mathbf{x}_k)$ from (4.25). Similarly, $L(\overline{\mathbf{x}}) = f(\overline{\mathbf{x}})$. Also, since $\overline{\mathbf{x}}$ is a KKT point, we have that $\nabla L(\overline{\mathbf{x}}) = \mathbf{0}$. Moreover, since $\mathbf{x}_k = \boldsymbol{\alpha}(\lambda_k)$ is feasible, $\mathbf{x}_k \to \overline{\mathbf{x}}$ as $\lambda_k \to 0^+$ or as $k \to \infty$, and since $\overline{\mathbf{x}}$ is a local minimum, we must have $f(\mathbf{x}_k) \ge f(\overline{\mathbf{x}})$ for k sufficiently large. Consequently, from (4.31), we get

$$\frac{f(\mathbf{x}_k) - f(\overline{\mathbf{x}})}{\lambda_k^2} = \frac{1}{2}\frac{(\mathbf{x}_k - \overline{\mathbf{x}})^t}{\lambda_k}\nabla^2 L(\overline{\mathbf{x}})\frac{\mathbf{x}_k - \overline{\mathbf{x}}}{\lambda_k}$$
$$+\left\|\frac{\mathbf{x}_k - \overline{\mathbf{x}}}{\lambda_k}\right\|^2 \beta[\overline{\mathbf{x}};(\mathbf{x}_k - \overline{\mathbf{x}}] \ge 0 \tag{4.32a}$$

for k large enough. But note that

$$\lim_{k\to\infty}\frac{\mathbf{x}_k - \overline{\mathbf{x}}}{\lambda_k} = \lim_{k\to\infty}\frac{\alpha(\lambda_k) - \alpha(0)}{\lambda_k} = \alpha'(0) = \mathrm{P}(0)\mathrm{d} = \mathrm{d} \tag{4.32b}$$

since d is already in the null space of the matrix having rows $\nabla g_i(\overline{\mathbf{x}})$ for $i \in I(\mathrm{d})$, and $\nabla h_i(\overline{\mathbf{x}})$ for $i = 1,\ldots, \ell$. Taking limits in (4.32a) as $k \to \infty$ and using (4.32b), we get that $\mathrm{d}^t\nabla^2 L(\overline{\mathbf{x}})\mathrm{d} \ge 0$, and this completes the proof.

Observe that the set C defined in the theorem is a subset of $G_0' \cap H_0$ and that the nonnegative curvature of L at $\overline{\mathbf{x}}$ is required for all $\mathrm{d} \in C$, but not necessarily for all $\mathrm{d} \in G_0' \cap H_0$. Furthermore, note that if the problem is unconstrained, Theorem 4.4.3 reduces to asserting that $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and $\mathbf{H}(\overline{\mathbf{x}})$ is positive semidefinite at a local minimum $\overline{\mathbf{x}}$. Hence, Theorem 4.1.3 is a special case of this result. Let us now illustrate the use of the foregoing results.

4.4.4 Example (McCormick [1967])

Consider the nonconvex programming problem

$$\text{P:} \quad \text{Minimize}\{(x_1-1)^2+x_2^2 : g_1(\mathbf{x}) = 2kx_1 - x_2^2 \le 0\},$$

where k is a positive constant. Figure 4.14 illustrates two possible ways in which the optimum is determined, depending on the value of k.

Note that $\nabla g_1(\mathbf{x}) = (2k, -2x_2)^t \neq (0,0)^t$, and hence the linear independence constraint qualification holds true at any feasible solution $\mathbf{x}$. The KKT conditions require primal feasibility and that

$$\begin{bmatrix} 2(x_1-1) \\ 2x_2 \end{bmatrix} + u_1 \begin{bmatrix} 2k \\ -2x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $u_1 \ge 0$ and $u_1[2kx_1 - x_2^2] = 0$. If $u_1 = 0$, we must have $(x_1, x_2) = (1, 0)$, which is the unconstrained minimum and which is infeasible for any $k > 0$. Hence, u_1 must be positive for any KKT point; so, by complementary slackness, $2kx_1 = x_2^2$ must hold true. Furthermore, by the second dual feasibility constraint, we must either have $x_2 = 0$ or $u_1 = 1$. If $x_2 = 0$, $2kx_1 = x_2^2$ yields $x_1 = 0$, and the first dual feasibility constraint yields $u_1 = 1/k$. This gives one KKT solution. Similarly, from the KKT conditions, when $u_1 = 1$, we obtain that $x_1 = 1 - k$ and $x_2 = \pm\sqrt{2k(1-k)}$, which yields a different set of KKT solutions when $0 < k < 1$. Hence, the KKT solutions are $\{\bar{\mathbf{x}}^1 = (0,0)^t, \ \bar{u}_1^1 = 1/k\}$ for any $k > 0$, and

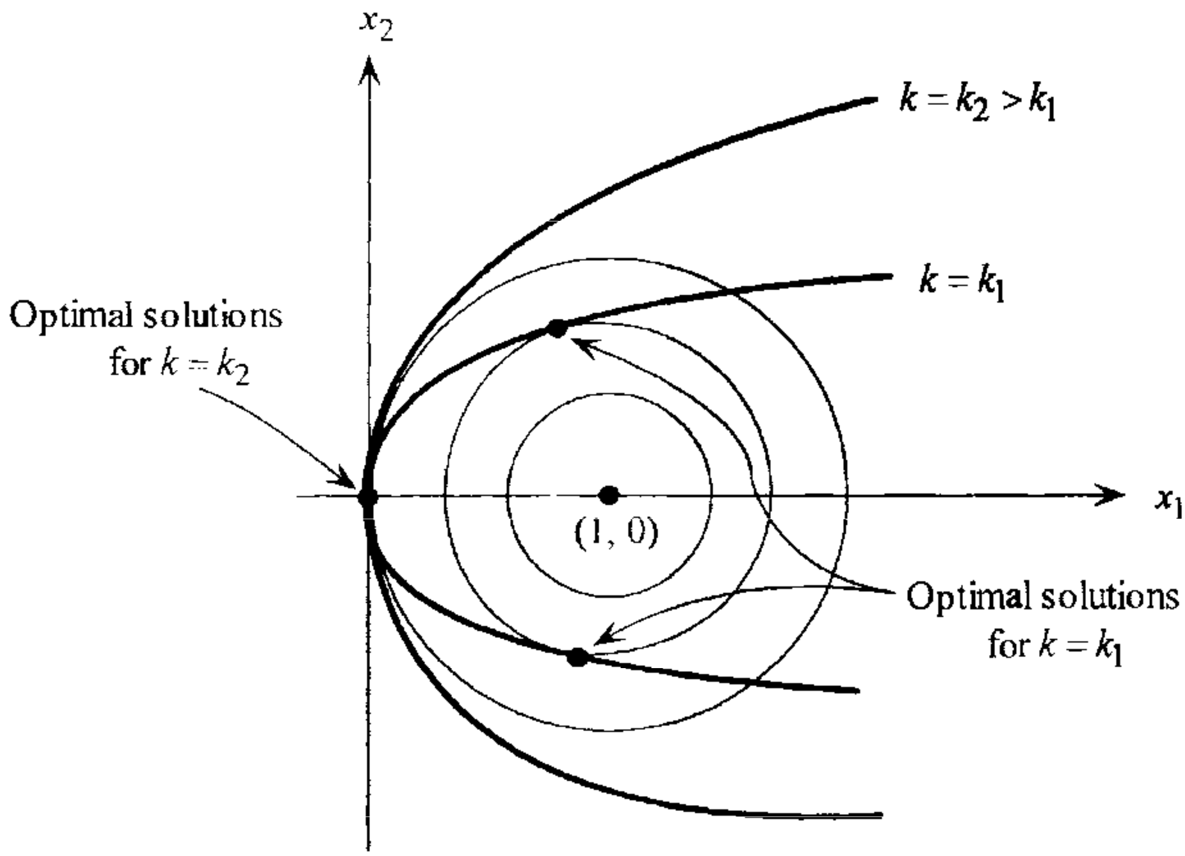


Figure 4.14 Two cases of optimal solutions: Example 4.4.4.

$\{\bar{\mathbf{x}}^2 = (1-k, \sqrt{2k(1-k)}, \ \bar{u}_1^2 = 1\}$ along with $\{\bar{\mathbf{x}}^3 = (1-k, -\sqrt{2k(1-k)}, \ \bar{u}_1^3 = 1\}$ whenever $0 < k < 1$.

By examining convex combinations of objective values at the above KKT points and at any other point $\mathbf{x}$ on the constraint surface, for example, it is readily verified that g_1 is not quasiconvex at these points; and thus, while the first-order necessary condition of Theorem 4.2.13 is satisfied, the sufficient condition of Theorem 4.2.16 does not hold true. Hence, we are uncertain about the character of the above KKT solutions using these results.

Now, let us examine the second-order necessary condition of Theorem 4.4.3. Note that $L(\mathbf{x}) = f(\mathbf{x}) + \bar{u}g(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + \bar{u}[2kx_1 - x_2^2]$, so

$$\nabla^2 L(\bar{\mathbf{x}}) = \begin{bmatrix} 2 & 0 \\ 0 & 2(1-\bar{u}) \end{bmatrix}.$$

Furthermore, the cone C defined in Theorem 4.4.3 is given by (since $\bar{u}_1 > 0$ at any KKT point)

$$C = \{\mathrm{d} \neq 0 : kd_1 = \bar{x}_2 d_2\}.$$

For the KKT solution $(\bar{\mathbf{x}}^1, \bar{u}_1^1)$, Theorem 4.4.3 requires that $2d_1^2 + 2(1-1/k)d_2^2 \geq 0$ for all (d_1, d_2) such that $d_1 = 0$. Whenever $k \geq 1$, this obviously holds true. However, when $0 < k < 1$, this condition is violated. Hence, using Theorem 4.4.3, we can conclude that $\bar{\mathbf{x}}^1$ is not a local minimum for $0 < k < 1$. On the other hand, since $\bar{u}_1^2 = \bar{u}_1^3 = 1$, $\nabla^2 L(\bar{\mathbf{x}}^2)$ and $\nabla^2 L(\bar{\mathbf{x}}^3)$ are positive semidefinite, and hence the other sets of KKT solutions satisfy the second-order necessary optimality conditions.

Next, let us examine the second-order sufficient conditions of Theorem 4.4.2. For the KKT solution $(\bar{\mathbf{x}}^1, \bar{u}_1^1)$, whenever $k > 1$, $\nabla^2 L(\bar{\mathbf{x}}^1)$ is itself positive definite; so, even by Lemma 4.4.1b, we have that $\bar{\mathbf{x}}^1$ is a strict local minimum. However, for $k = 1$, although $\bar{\mathbf{x}}^1$ solves Problem P, we are unable to recognize this via Theorem 4.4.2, since $\mathbf{d}^t \nabla^2 L(\bar{\mathbf{x}}^1)\mathbf{d} = 2d_1^2 = 0$ for $\mathbf{d} \in C = \{\mathbf{d} \neq 0 : d_1 = 0\}$.

Next, consider the KKT solution $(\bar{\mathbf{x}}^2, \bar{u}_1^2)$ for $0 < k < 1$. Here $C = \{\mathbf{d} \neq 0 : kd_1 = \sqrt{2k(1-k)}d_2\}$; and for any $\mathbf{d}$ in C, we have $\mathbf{d}^t \nabla^2 L(\bar{\mathbf{x}}^2)\mathbf{d} = 2d_1^2 > 0$. Hence, by Theorem 4.4.2, $\bar{\mathbf{x}}^2$ is a strict local minimum for $0 < k < 1$. Note that $\nabla^2 L(\bar{\mathbf{x}}^2)$ itself is not positive definite, and therefore Theorem 4.4.2 plays a critical role in concluding the local minimum status of $\bar{\mathbf{x}}^2$. Similarly, $\bar{\mathbf{x}}^3$ is a strict local minimum for $0 < k < 1$. The global minimum status of the strict local minima above must be established by other means because of the nonconvexity of the problem (see Exercise 4.40).

Exercises

[4.1] Consider the univariate function $f(x) = xe^{-2x}$. Find all local minima/maxima and inflection points. Also, what can you claim about a global minimum and a global maximum for f? Give analytical justifications for your claims.

[4.2] Consider the following linear program:

$$\begin{aligned} \text{Maximize} \quad & x_1 + 3x_2 \\ \text{subject to} \quad & 2x_1 + 3x_2 \le 6 \\ & -x_1 + 4x_2 \le 4 \\ & x_1, x_2 \ge 0. \end{aligned}$$

a. Write the KKT optimality conditions.
b. For each extreme point, verify whether or not the KKT conditions hold true, both algebraically and geometrically. From this, find an optimal solution.

[4.3] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & x_1^2 + 2x_2^2 \\ \text{subject to} \quad & x_1 + x_2 - 2 = 0. \end{aligned}$$

Find a point satisfying the KKT conditions and verify that it is indeed an optimal solution. Re-solve the problem if the objective function is replaced by $x_1^3 + x_2^3$.

[4.4] Consider the following unconstrained problem:

$$\text{Minimize } 2x_1^2 - x_1x_2 + x_2^2 - 3x_1 + e^{2x_1+x_2}.$$

a. Write the first-order necessary optimality conditions. Is this condition also sufficient for optimality? Why?
b. Is $\bar{\mathbf{x}} = (0,0)^t$ an optimal solution? If not, identify a direction d along which the function would decrease.
c. Minimize the function starting from (0, 0) along the direction d obtained in Part b.
d. Dropping the last term in the objective function, use a classical direct optimization technique to solve this problem.

[4.5] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & x_1^4 + x_2^4 + 12x_1^2 + 6x_2^2 - x_1x_2 - x_1 - x_2 \\ \text{subject to} \quad & x_1 + x_2 \ge 6 \\ & 2x_1 - x_2 \ge 3 \\ & x_1 \ge 0,\ x_2 \ge 0. \end{aligned}$$

Write out the KKT conditions and show that $(x_1, x_2) = (3, 3)$ is the unique optimal solution.

[4.6] Consider the problem to minimize $\|\mathbf{Ax} - \mathbf{b}\|^2$, where **A** is an $m \times n$ matrix and **b** is an m-vector.

a. Give a geometric interpretation of the problem.
b. Write a necessary condition for optimality. Is this also a sufficient condition?
c. Is the optimal solution unique? Why or why not?
d. Can you give a closed-form solution of the optimal solution? Specify any assumptions that you may need.
e. Solve the problem for

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 6 \\ 2 \\ 0 \end{bmatrix}.$$

[4.7] Consider the following problem:

$$\begin{aligned} \text{Minimize} \ & \left(x_1 - \frac{9}{4}\right)^2 + (x_2 - 2)^2 \\ \text{subject to} \ & x_2 - x_1^2 \geq 0 \\ & x_1 + x_2 \leq 6 \\ & x_1, \quad x_2 \geq 0. \end{aligned}$$

a. Write the KKT optimality conditions and verify that these conditions hold true at the point $\bar{\mathbf{x}} = (3/2, 9/4)^t$.
b. Interpret the KKT conditions at $\bar{\mathbf{x}}$ graphically.
c. Show that $\bar{\mathbf{x}}$ is indeed the unique global optimal solution.

[4.8] Consider the following problem:

$$\begin{aligned} \text{Minimize} \ & \frac{x_1 + 3x_2 + 3}{2x_1 + x_2 + 6} \\ \text{subject to} \ & 2x_1 + x_2 \leq 12 \\ & -x_1 + 2x_2 \leq 4 \\ & x_1, \quad x_2 \geq 0. \end{aligned}$$

a. Show that the KKT conditions are sufficient for this problem.
b. Show that any point on the line segment joining the points (0, 0) and (6, 0) is an optimal solution.

[4.9] Consider the following problem, where **c** is a nonzero vector in R^n:

$$\text{Maximize} \quad \mathbf{c}^t\mathbf{d}$$
$$\text{subject to} \quad \mathbf{d}^t\mathbf{d} \le 1.$$

a. Show that $\bar{\mathbf{d}} = \mathbf{c}/\|\mathbf{c}\|$ is a KKT point. Furthermore, show that $\bar{\mathbf{d}}$ is indeed the unique global optimal solution.
b. Using the result of Part a, show that the direction of steepest ascent of f at a point $\mathbf{x}$ is given by $\nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$ provided that $\nabla f(\mathbf{x}) \neq \mathbf{0}$.

[4.10] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$.

a. Show that verifying whether a point $\bar{\mathbf{x}}$ is a KKT point is equivalent to finding a vector $\mathbf{u}$ satisfying a system of the form $\mathbf{A}^t\mathbf{u} = \mathbf{c}$, $\mathbf{u} \ge \mathbf{0}$. (This can be done using Phase I of linear programming.)
b. Indicate the modifications needed in Part a if the problem had equality constraints.
c. Illustrate Part a by the following problem, where $\bar{\mathbf{x}} = (1,2,5)^t$:

$$\begin{array}{ll}
\text{Minimize} & 2x_1^2 + x_2^2 + 2x_3^2 + x_1x_3 - x_1x_2 + x_1 + 2x_3 \\
\text{subject to} & x_1^2 + x_2^2 - x_3 \le 0 \\
& x_1 + x_2 + 2x_3 \le 16 \\
& x_1 + x_2 \ge 3 \\
& x_1, \quad x_2, \quad x_3 \ge 0.
\end{array}$$

[4.11] Consider the problem to minimize $f(x_1,x_2) = (x_2^2 - x_1)(x_2^2 - 2x_1)$, and let $\bar{\mathbf{x}} = (0,0)^t$. Show that for each $\mathbf{d} \in R^n$, $\|\mathbf{d}\| = 1$, there exists a $\delta_{\mathbf{d}} > 0$ such that for $-\delta_{\mathbf{d}} \le \lambda \le \delta_{\mathbf{d}}$, we have $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) \ge f(\bar{\mathbf{x}})$. However, show that $\inf\{\delta_{\mathbf{d}} : \mathbf{d} \in R^n, \|\mathbf{d}\| = 1\} = 0$. In reference to Figure 4.1, discuss what this entails regarding the local optimality of $\bar{\mathbf{x}}$.

[4.12] Consider the following problem, where a_j, b, and c_j are positive constants:

$$\begin{array}{ll}
\text{Minimize} & \displaystyle\sum_{j=1}^{n} \frac{c_j}{x_j} \\
\text{subject to} & \displaystyle\sum_{j=1}^{n} a_j x_j = b \\
& x_j \ge 0 \qquad \text{for } j = 1,\ldots,n.
\end{array}$$

Write the KKT conditions and solve for the point $\bar{\mathbf{x}}$ satisfying these conditions.

[4.13] Consider Problem P to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$. Suppose that this problem is reformulated as $\overline{\text{P}}$:

Minimize $\{f(\mathbf{x}): g_i(\mathbf{x})+s_i^2 = 0$ for $i = 1,\ldots, m$ and $h_i(\mathbf{x})=0$ for $i = 1,\ldots, \ell\}$. Write the KKT conditions for P and for $\overline{\text{P}}$ and compare them. Explain any difference between the two and what arguments you can use to resolve them. Express your opinion on using the formulation $\overline{\text{P}}$ to solve the problem.

[4.14] Consider Problem P to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$. Show that P is mathematically equivalent to the following single constraint problem $\overline{\text{P}}$, where $s_1,\ldots,s_m$ are additional variables:

$$\overline{\text{P}}: \text{ Minimize}\left\{ f(\mathbf{x}): \sum_{i=1}^{m}[g_i(\mathbf{x})+s_i^2]^2 + \sum_{i=1}^{\ell} h_i^2(\mathbf{x}) = 0 \right\}.$$

Write out the FJ and the KKT conditions for $\overline{\text{P}}$. What statements can you make about the relationship between local optima, FJ, and KKT points? What is your opinion about the utility of $\overline{\text{P}}$ in solving P?

[4.15] In geometric programming, the following result is used. If $x_1,\ldots,x_n \ge 0$,

$$\frac{1}{n}\sum_{j=1}^{n} x_j \ge \left(\prod_{j=1}^{n} x_j\right)^{1/n}.$$

Prove the result using the KKT conditions.

[*Hint*: Consider one of the following problems and justify your use of it:

- Minimize $\sum_{j=1}^{n} x_j$

  subject to $\prod_{j=1}^{n} x_j = 1, x_j \ge 0$ for $j = 1,\ldots,n$.

- Maximize $\prod_{j=1}^{n} x_j$

  subject to $\sum_{j=1}^{n} x_j = 1, x_j \ge 0$ for $j = 1,\ldots,n$.]

[4.16] Consider the quadratic assignment program to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{Q}\mathbf{x}$ subject to the assignment constraints $\sum_{j=1}^{m} x_{ij} = 1$ for all $i = 1,\ldots, m$, $\sum_{i=1}^{m} x_{ij} = 1$ for all $j = 1,\ldots, m$, and $\mathbf{x}$ is binary valued. Here the component x_{ij} of the vector $\mathbf{x}$ takes on a value of 1 if i is assigned to j and is 0 otherwise, for $i, j = 1,\ldots, m$. Show that if M exceeds the sum of absolute values of elements in any row of $\mathbf{Q}$, the matrix $\overline{\mathbf{Q}}$ obtained from $\mathbf{Q}$ by subtracting M from each diagonal element is negative definite. Now, consider the following problem:

$$\overline{\text{QAP}}: \text{ Minimize}\left\{\mathbf{c}^t\mathbf{x}+\frac{1}{2}\mathbf{x}^t\overline{\mathbf{Q}}\mathbf{x}: \sum_{j=1}^{m} x_{ij}=1 \text{ for all } i, \sum_{i=1}^{m} x_{ij}=1 \text{ for all } j, \mathbf{x}\geq \mathbf{0}\right\}.$$

Using the fact that the extreme points of $\overline{\text{QAP}}$ are all binary valued, show that $\overline{\text{QAP}}$ is equivalent to QAP. Moreover, show that every extreme point of $\overline{\text{QAP}}$ is a KKT point. (This exercise is due to Bazaraa and Sherali [1982].)

[4.17] Answer the following and justify your answer:

a. For a minimization nonlinear program, can a KKT point be a local maximum?
b. Let f be differentiable, let X be convex, and let $\overline{\mathbf{x}} \in X$ satisfy $\nabla f(\overline{\mathbf{x}})^t(\mathbf{x}-\overline{\mathbf{x}}) > 0$ for all $\mathbf{x} \in X$, $\mathbf{x} \neq \overline{\mathbf{x}}$. Is $\overline{\mathbf{x}}$ necessarily a local minimum?
c. What is the effect on application of the FJ and KKT optimality conditions by duplicating an equality constraint or an inequality constraint in the problem?

[4.18] Write the KKT necessary optimality conditions for Exercises 1.3 and 1.4. Using these conditions, find the optimal solutions.

[4.19] Let $f: R^n \to R$ be infinitely differentiable, and let $\overline{\mathbf{x}} \in R^n$. For any $\mathbf{d} \in R^n$, define $F_{\mathbf{d}}(\lambda) = f(\overline{\mathbf{x}} + \lambda \mathbf{d})$ for $\lambda \in R$. Write out the infinite Taylor series expansion for $F_{\mathbf{d}}(\lambda)$ and compute $F''_{\mathbf{d}}(\lambda)$. Compare the nonnegativity or positivity of this expression with the necessary and sufficient Taylor series–based inequality for $\overline{\mathbf{x}}$ to be a local minimum for f. What conclusions can you draw?

[4.20] Consider the following one-dimensional minimization problem:

$$\begin{aligned} &\text{Minimize} \quad f(\mathbf{x}+\lambda\mathbf{d}) \\ &\text{subject to} \quad \lambda \geq 0, \end{aligned}$$

where $\mathbf{x}$ is a given vector and $\mathbf{d}$ is a given nonzero direction.

a. Write a necessary condition for a minimum if f is differentiable. Is this condition also sufficient? If not, what assumptions on f would make the necessary condition also sufficient?
b. Suppose that f is convex but not differentiable. Can you develop a necessary optimality condition for the above problem using subgradients of f as defined in Section 3.2?

[4.21] Use the KKT conditions to prove Farkas's theorem discussed in Section 2.3. (*Hint*: Consider the problem to maximize $\mathbf{c}^t\mathbf{x}$ subject to $\mathbf{Ax} \leq \mathbf{0}$.)

[4.22] Suppose that $f: S \to R$, where $S \subseteq R^n$.

a. If f is pseudoconvex over $N_\varepsilon(\overline{\mathbf{x}}) \cap S$, does this imply that f is pseudoconvex at $\overline{\mathbf{x}}$?

b. If f is strictly pseudoconvex at $\bar{\mathbf{x}}$, does this imply that f is quasiconvex at $\bar{\mathbf{x}}$?

c. For each of the FJ and KKT sufficiency theorems for both the equality and equality–inequality constraint cases, provide alternative sets of sufficient conditions to guarantee local optimality of a point satisfying these conditions. Prove your claims. Examine your proof for possibly strengthening the theorem by weakening your assumptions.

[4.23] Let X be a nonempty open set in R^n, and consider $f: R^n \to R$, $g_i: R^n \to R$ for $i = 1,\ldots, m$, and $h_i: R^n \to R$, for $i = 1,\ldots, \ell$. Consider Problem P:

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) \\ &\text{subject to} && g_i(\mathbf{x}) \le 0 && \text{for } i = 1,\ldots,m \\ &&& h_i(\mathbf{x}) = 0 && \text{for } i = 1,\ldots,\ell \\ &&& \mathbf{x} \in X. \end{aligned}$$

Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that the KKT conditions hold true at $\bar{\mathbf{x}}$; that is, there exist scalars $\bar{u}_i \ge 0$ for $i \in I$ and $\bar{v}_i$ for $i = 1,\ldots,\ell$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i\in I} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

a. Suppose that f is pseudoconvex at $\bar{\mathbf{x}}$ and that ϕ is quasiconvex at $\bar{\mathbf{x}}$, where

$$\phi(\mathbf{x}) = \sum_{i\in I} \bar{u}_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} \bar{v}_i h_i(\mathbf{x}).$$

Show that $\bar{\mathbf{x}}$ is a global optimal solution to Problem P.

b. Show that if $f + \sum_{i\in I} \bar{u}_i g_i + \sum_{i=1}^{\ell} \bar{v}_i h_i$ is pseudoconvex, $\bar{\mathbf{x}}$ is a global optimal solution to Problem P.

c. Show by means of examples that the convexity assumptions in Parts a and b and those of Theorem 4.3.8 are not equivalent to each other.

d. Relate this result to Lemma 4.4.1 and to the discussion immediately following it.

[4.24] Let $\bar{\mathbf{x}}$ be an optimal solution to the problem of minimizing $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$, $i = 1,\ldots, m$ and $h_i(\mathbf{x}) = 0$, $i = 1,\ldots, \ell$. Suppose that $g_k(\bar{\mathbf{x}}) < 0$ for some $k \in \{1,\ldots, m\}$. Show that if this nonbinding constraint is deleted, it is possible that $\bar{\mathbf{x}}$ is not even a local minimum for the resulting problem. [*Hint*: Consider $g_k(\bar{\mathbf{x}}) = -1$ and $g_k(\mathbf{x}) = 1$ for $\mathbf{x} \neq \bar{\mathbf{x}}$.] Show that if all problem-defining functions are continuous, then, by deleting nonbinding constraints, $\bar{\mathbf{x}}$ remains at least a local optimal solution.

[4.25] Consider the *bilinear program* to minimize $\mathbf{c}^t\mathbf{x} + \mathbf{d}^t\mathbf{y} + \mathbf{x}^t\mathbf{H}\mathbf{y}$ subject to $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, where X and Y are bounded polyhedral sets in R^n and R^m, respectively. Let $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ be extreme points of the sets X and Y, respectively.

a. Verify that the objective function is neither quasiconvex nor quasiconcave.
b. Prove that there exists an extreme point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ that solves the bilinear program.
c. Prove that the point $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is a local minimum of the bilinear program if and only if the following are true: (i) $\mathbf{c}^t(\mathbf{x}-\hat{\mathbf{x}}) \ge 0$ and $\mathbf{d}^t(\mathbf{y}-\hat{\mathbf{y}}) \ge 0$ for each $\mathbf{x} \in X$ and $\mathbf{y} \in Y$; (ii) $\mathbf{c}^t(\mathbf{x}-\hat{\mathbf{x}}) + \mathbf{d}^t(\mathbf{y}-\hat{\mathbf{y}}) > 0$ whenever $(\mathbf{x}-\hat{\mathbf{x}})^t\mathbf{H}(\mathbf{y}-\hat{\mathbf{y}}) < 0$.
d. Show that the point $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is a KKT point if and only if $(\mathbf{c}^t + \hat{\mathbf{y}}^t\mathbf{H})(\mathbf{x}-\hat{\mathbf{x}}) \ge 0$ for each $\mathbf{x} \in X$ and $(\mathbf{d}^t + \hat{\mathbf{x}}^t\mathbf{H})(\mathbf{y}-\hat{\mathbf{y}}) \ge 0$ for each $\mathbf{y} \in Y$.
e. Consider the problem to minimize $x_2 + y_1 + x_2y_1 - x_1y_2 + x_2y_2$ subject to $(x_1, x_2) \in X$ and $(y_1, y_2) \in Y$, where X is the polyhedral set defined by its extreme points (0, 0), (0, 1), (1, 4), (2, 4), and (3, 0), and Y is the polyhedral set defined by its extreme points (0, 0), (0, 1), (1, 5), (3, 5), (4, 4), and (3, 0). Verify that the point $(x_1, x_2, y_1, y_2) = (0, 0, 0, 0)$ is a KKT point but not a local minimum. Verify that the point $(x_1, x_2, y_1, y_2) = (3, 0, 1, 5)$ is both a KKT point and a local minimum. What is the global minimum to the problem?

[4.26] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \ge \mathbf{0}$, where f is a differentiable convex function. Let $\bar{\mathbf{x}}$ be a given point and denote $\nabla f(\bar{\mathbf{x}})$ by $(\nabla_1, \ldots, \nabla_n)^t$. Show that $\bar{\mathbf{x}}$ is an optimal solution if and only if $\mathbf{d} = \mathbf{0}$, where $\mathbf{d}$ is defined by

$$d_i = \begin{cases} -\nabla_i & \text{if } x_i > 0 \text{ or } \nabla_i < 0 \\ 0 & \text{if } x_i = 0 \text{ and } \nabla_i \ge 0. \end{cases}$$

[4.27] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I = \{i: g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f is differentiable at $\bar{\mathbf{x}}$ and that each g_i for $i \in I$ is differentiable and concave at $\bar{\mathbf{x}}$. Furthermore, suppose that each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$. Consider the following linear program:

$$\begin{aligned} &\text{Minimize} && \nabla f(\bar{\mathbf{x}})^t \mathbf{d} && \\ &\text{subject to} && \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 && \text{for } i \in I \\ & && -1 \le d_j \le 1 && \text{for } j = 1,\dots,n. \end{aligned}$$

Let $\bar{\mathbf{d}}$ be an optimal solution with objective function value $\bar{z}$.

a. Show that $\bar{z} \le 0$.
b. Show that if $\bar{z} < 0$, there exists a $\delta > 0$ such that $\bar{\mathbf{x}} + \lambda\bar{\mathbf{d}}$ is feasible and $f(\bar{\mathbf{x}} + \lambda\bar{\mathbf{d}}) < f(\bar{\mathbf{x}})$ for each $\lambda \in (0, \delta)$.
c. Show that if $\bar{z} = 0$, $\bar{\mathbf{x}}$ satisfies the KKT conditions.

[4.28] Consider the following problem, where $\mathbf{y}$, $\mathbf{e}$, and $\mathbf{y}_0$ belong to R^n, and where $\mathbf{y} = (y_1,\dots,y_n)^t$, $\mathbf{y}_0 = (1/n,\dots,1/n)^t$, and $\mathbf{e} = (1,\dots,1)^t$:

$$\text{Minimize } \{y_1 : \|\mathbf{y} - \mathbf{y}_0\|^2 \le 1/n(n-1),\ \mathbf{e}^t\mathbf{y} = 1\}.$$

Interpret this problem with respect to an inscribed sphere in the simplex defined by $\{\mathbf{y} : \mathbf{e}^t\mathbf{y} = 1,\ \mathbf{y} \ge \mathbf{0}\}$. Write the KKT conditions for this problem and verify that $(0, 1/(n-1),\dots,1/(n-1))$ is an optimal solution.

[4.29] Let f: $R^n \to R$, g_i: $R^n \to R$ for $i = 1,\dots, m$ be convex functions. Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\dots, m$. Let M be a proper subset of $\{1,\dots, m\}$, and suppose that $\hat{\mathbf{x}}$ solves the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i \in M$. Let $V = \{i : g_i(\hat{\mathbf{x}}) > 0\}$. If $\bar{\mathbf{x}}$ solves the original problem, and if $f(\bar{\mathbf{x}}) > f(\hat{\mathbf{x}})$, show that $g_i(\bar{\mathbf{x}}) = 0$ for some $i \in V$. Show that this is not necessarily true if $f(\bar{\mathbf{x}}) = f(\hat{\mathbf{x}})$. (This exercise also shows that if an unconstrained minimum of f is infeasible and has an objective value less than the optimum value, any constrained minimum lies on the boundary of the feasible region.)

[4.30] Consider Problem P, to minimize $f(\mathbf{x})$ subject to some "less than or equal to" type of linear inequality constraints. Let $\bar{\mathbf{x}}$ be a feasible solution, and let the binding constraints be represented as $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix of rank m. Let $\mathbf{d} = -\nabla f(\bar{\mathbf{x}})$ and consider the following problem.

$$\bar{\text{P}}: \ \text{Minimize}\left\{\frac{1}{2}\|\mathbf{x} - (\bar{\mathbf{x}} + \mathbf{d})\|^2 : \mathbf{A}\mathbf{x} = \mathbf{b}\right\}.$$

Let $\hat{\mathbf{x}}$ solve $\bar{\text{P}}$.

a. Provide a geometric interpretation for $\bar{\text{P}}$ and its solution $\hat{\mathbf{x}}$.
b. Write the KKT conditions for $\bar{\text{P}}$. Discuss whether these conditions are necessary and sufficient for optimality.

c. Suppose that the given point $\bar{\mathbf{x}}$ happens to be a KKT point for $\overline{\mathrm{P}}$. Is $\bar{\mathbf{x}}$ also a KKT point for P? If so, why? If not, under what additional conditions can you make this claim?

d. Determine a closed-form expression for the solution $\hat{\mathbf{x}}$ to Problem $\overline{\mathrm{P}}$.

[4.31] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Let $\bar{\mathbf{x}}^t = (\bar{\mathbf{x}}_B^t, \bar{\mathbf{x}}_N^t)$ be an extreme point, where $\bar{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$, $\bar{\mathbf{x}}_N = \mathbf{0}$, and $\mathbf{A} = [\mathbf{B}, \mathbf{N}]$ with $\mathbf{B}$ invertible. Now, consider the following direction-finding problem:

$$\begin{aligned} \text{Minimize} \quad & [\nabla_N f(\bar{\mathbf{x}}) - \nabla_B f(\bar{\mathbf{x}})\mathbf{B}^{-1}\mathbf{N}]^t \mathbf{d}_N \\ \text{subject to} \quad & 0 \leq d_j \leq 1 \qquad \text{for each nonbasic component } j, \end{aligned}$$

where $\nabla_B f(\bar{\mathbf{x}})$ and $\nabla_N f(\bar{\mathbf{x}})$ denote the gradient of f with respect to the basic and nonbasic variables, respectively. Let $\bar{\mathbf{d}}_N$ be an optimal solution, and let $\bar{\mathbf{d}}_B = -\mathbf{B}^{-1}\mathbf{N}\bar{\mathbf{d}}_N$. Show that if $\bar{\mathbf{d}}^t = (\bar{\mathbf{d}}_B^t, \bar{\mathbf{d}}_N^t) \neq (\mathbf{0}, \mathbf{0})$, it is an improving feasible direction. What are the implications of $\bar{\mathbf{d}} = \mathbf{0}$?

[4.32] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & \sum_{j=1}^{n} f_j(x_j) \\ \text{subject to} \quad & \sum_{j=1}^{n} x_j = 1 \\ & x_j \geq 0 \qquad \text{for } j = 1,\dots,n. \end{aligned}$$

Suppose that $\bar{\mathbf{x}} = (\bar{x}_1,\dots,\bar{x}_n)^t \geq \mathbf{0}$ solves the problem. Letting $\delta_j = \partial f_j(\bar{\mathbf{x}})/\partial x_j$, show that there exists a scalar k such that

$$\delta_j \geq k \qquad \text{and} \qquad (\delta_j - k)\bar{x}_j = 0 \qquad \text{for } j = 1,\dots,n.$$

[4.33] Let $\mathbf{c}$ be an n-vector, $\mathbf{b}$ an m-vector, $\mathbf{A}$ an $m \times n$ matrix, and $\mathbf{H}$ a symmetric $n \times n$ positive definite matrix. Consider the following two problems:

- Minimize $\mathbf{c}^t\mathbf{x}+\frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x}$

  subject to $\mathbf{A}\mathbf{x}\leq\mathbf{b}$;
- Minimize $\mathbf{h}^t\mathbf{v}+\frac{1}{2}\mathbf{v}^t\mathbf{G}\mathbf{v}$

  subject to $\mathbf{v}\geq\mathbf{0}$,

where $\mathbf{G}=\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^t$ and $\mathbf{h}=\mathbf{A}\mathbf{H}^{-1}\mathbf{c}+\mathbf{b}$. Investigate the relationship between the KKT conditions of these two problems.

[4.34] Consider the following problem:

$$\begin{aligned} &\text{Minimize} && -x_1+x_2 \\ &\text{subject to} && x_1^2+x_2^2-2x_1=0 \\ &&& (x_1,x_2)\in X, \end{aligned}$$

where X is the convex combinations of the points $(-1,0)$, $(0,1)$, $(1,0)$, and $(0,-1)$.

a. Find the optimal solution graphically.
b. Do the Fritz John or KKT conditions hold at the optimal solution in Part a? If not, explain in terms of Theorems 4.3.2 and 4.3.7.
c. Replace the set X by a suitable system of inequalities and answer Part b. What are your conclusions?

[4.35] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$, where $f\colon R^n\to R$ and $g_i\colon R^n\to R$ for $i=1,\ldots,m$, are all differentiable functions. We know that if $\bar{\mathbf{x}}$ is a local minimum, then $F\cap D=\varnothing$, where F and D are, respectively, the set of improving and feasible directions. Show, giving examples, that the converse is false even if f is convex or if the feasible region is convex (although not both). However, suppose that there exists an ε-neighborhood $N_\varepsilon(\bar{\mathbf{x}})$ about $\bar{\mathbf{x}}$, $\varepsilon>0$, such that f is pseudoconvex and g_i for $i\in I=\{i : g_i(\bar{\mathbf{x}})=0\}$ are quasiconvex over $N_\varepsilon(\bar{\mathbf{x}})$. Show that $\bar{\mathbf{x}}$ is a local minimum if and only if $F\cap D=\varnothing$. (*Hint*: Examine Lemma 4.2.3 and Theorem 4.2.5.) Extend this result to include equality constraints.

[4.36] Consider the following problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x})\leq 0$ for $i=1,\ldots,m$ and $h_i(\mathbf{x})=0$ for $i=1,\ldots,\ell$. Suppose that $\bar{\mathbf{x}}$ solves the problem locally, and let $I=\{i\colon g_i(\bar{\mathbf{x}})=0\}$. Furthermore, suppose that each g_i for $i\in I$ is differentiable at $\bar{\mathbf{x}}$, each g_i for $i\notin I$ is continuous at $\bar{\mathbf{x}}$, and $h_1,\ldots,h_\ell$ are affine; that is, each h_i is of the form $h_i(\mathbf{x})=\mathbf{a}_i^t\mathbf{x}-b_i$.

a. Show that $F_0\cap G\cap H_0=\varnothing$, where

   $F_0=\{\mathrm{d} : \nabla f(\bar{\mathbf{x}})^t\mathbf{d}<0\}$

$H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\ldots, \ell\}$

$G = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for } i \in J \text{ and } \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0 \text{ for } i \in I - J\}$,

where $J = \{i \in I : g_i \text{ is pseudoconcave at } \bar{\mathbf{x}}\}$.

b. Show how this condition can be verified by using linear programming.

c. Dropping the equality constraints $h_i(\bar{\mathbf{x}}) = 0$, $i = 1,\ldots, \ell$, from the problem and letting D denote the resulting set of feasible directions, show that $G \subseteq D$ and hence that $F_0 \cap G = \varnothing$.

[4.37] Consider the following problem:

$$\begin{aligned} &\text{Maximize } x_1^2 + 4x_1x_2 + x_2^2 \\ &\text{subject to } x_1^2 + x_2^2 = 1. \end{aligned}$$

a. Using the KKT conditions, find an optimal solution to the problem.
b. Test for the second-order optimality conditions.
c. Does the problem have a unique optimal solution?

[4.38] Consider the following problem:

$$\begin{aligned} \text{Maximize } & 3x_1 - x_2 + x_2^3 \\ \text{subject to } & x_1 + x_2 + x_3 \le 0 \\ & -x_1 + 2x_2 + x_3^2 = 0. \end{aligned}$$

a. Write the KKT optimality conditions.
b. Test for the second-order optimality conditions.
c. Argue why this problem is unbounded.

[4.39] Consider the following problem:

$$\begin{aligned} \text{Maximize } & (x_1 - 2)^2 + (x_2 - 3)^2 \\ \text{subject to } & 3x_1 + 2x_2 \ge 6 \\ & -x_1 + x_2 \le 3 \\ & x_1 \le 2. \end{aligned}$$

a. Graphically, find all locally maximizing solutions. What is the global maximum for this problem?
b. Repeat Part a analytically, using first- and second-order KKT optimality conditions along with any other formal optimality characterizations.

[4.40] Consider the Problem of Example 4.4.4 for the case $k = 1$. Provide an analytical argument to show that $\bar{\mathbf{x}} = (0,0)^t$ is an optimal solution. By examining a sequence of values of $k \to 1^-$ with respect to the point (0, 0), explain why the second-order optimality conditions are unable to resolve this case.

[4.41] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{Ax} \le \mathbf{b}$. Suppose that $\bar{\mathbf{x}}$ is a feasible solution such that $\mathbf{A}_1\bar{\mathbf{x}} = \mathbf{b}_1$ and $\mathbf{A}_2\bar{\mathbf{x}} < \mathbf{b}_2$, where $\mathbf{A}^t = (\mathbf{A}_1^t, \mathbf{A}_2^t)$ and $\mathbf{b}^t = (\mathbf{b}_1^t, \mathbf{b}_2^t)$. Assuming that $\mathbf{A}_1$ has full rank, the matrix $\mathbf{P}$ that projects any vector unto the nullspace of $\mathbf{A}_1$ is given by

$$\mathbf{P} = \mathbf{I} - \mathbf{A}_1^t(\mathbf{A}_1\mathbf{A}_1^t)^{-1}\mathbf{A}_1.$$

a. Let $\bar{\mathbf{d}} = -\mathbf{P}\nabla f(\bar{\mathbf{x}})$. Show that if $\bar{\mathbf{d}} \ne 0$, it is an improving feasible direction; that is, $\bar{\mathbf{x}} + \lambda\mathbf{d}$ is feasible and $f(\bar{\mathbf{x}} + \lambda\mathbf{d}) < f(\bar{\mathbf{x}})$ for $\lambda > 0$ and sufficiently small.

b. Suppose that $\bar{\mathbf{d}} = \mathbf{0}$ and that $\mathbf{u} = -(\mathbf{A}_1\mathbf{A}_1^t)^{-1}\mathbf{A}_1\nabla f(\bar{\mathbf{x}}) \ge \mathbf{0}$. Show that $\bar{\mathbf{x}}$ is a KKT point.

c. Show that $\bar{\mathbf{d}}$ generated above is of the form $\lambda\hat{\mathbf{d}}$ for some $\lambda > 0$, where $\hat{\mathbf{d}}$ is an optimal solution to the following problem:

$$\begin{aligned} \text{Minimize} \quad & \nabla f(\bar{\mathbf{x}})^t\mathbf{d} \\ \text{subject to} \quad & \mathbf{A}_1\mathbf{d} = \mathbf{0} \\ & \|\mathbf{d}\|^2 \le 1. \end{aligned}$$

d. Make all possible simplifications if $\mathbf{A} = -\mathbf{I}$ and $\mathbf{b} = \mathbf{0}$, that is, if the constraints are of the form $\mathbf{x} \ge \mathbf{0}$.

[4.42] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & x_1^2 - x_1x_2 + 2x_2^2 - 4x_1 - 5x_2 \\ \text{subject to} \quad & x_1 + 2x_2 \le 6 \\ & x_1 \qquad\ \ \le 2 \\ & x_1, \quad x_2 \ge 0. \end{aligned}$$

a. Solve the problem geometrically and verify the optimality of the solution obtained by the KKT conditions.

b. Find the direction $\bar{\mathbf{d}}$ of Exercise 4.41 at the optimal solution. Verify that $\bar{\mathbf{d}} = \mathbf{0}$ and that $\mathbf{u} \ge \mathbf{0}$.

c. Find the direction $\bar{\mathbf{d}}$ of Exercise 4.41 at $\bar{\mathbf{x}} = (1, 5/2)^t$. Verify that $\bar{\mathbf{d}}$ is an improving feasible direction. Also, verify that the optimal solution $\hat{\mathbf{d}}$ of Part c of Exercise 4.41 indeed points along $\bar{\mathbf{d}}$.

[4.43] Let A be an $m \times n$ matrix of rank m, and let $\mathbf{P} = \mathbf{I} - \mathbf{A}^t(\mathbf{AA}^t)^{-1}\mathbf{A}$ be the matrix that projects any vector onto the null space of $\mathbf{A}$. Define $C = \{\mathbf{d} : \mathbf{Ad} = \mathbf{0}\}$, and let H be an $n \times n$ symmetric matrix. Show that $\mathbf{d} \in C$ if and only if $\mathbf{d} =$

$\mathbf{Pw}$ for some $\mathbf{w} \in R^n$. Show that $\mathbf{d}^t\mathbf{Hd} \geq 0$ for all $\mathbf{d} \in C$ if and only if $\mathbf{P}^t\mathbf{HP}$ is positive semidefinite.

[4.44] Consider Problem P to minimize $f(\mathbf{x})$ subject to $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$. Let $\bar{\mathbf{x}}$ be a feasible solution and define $\mathbf{A}$ as an $\ell \times n$ matrix whose rows represent $\nabla h_i(\bar{\mathbf{x}})^t$ for $i = 1,\ldots, \ell$ and assume that $\mathbf{A}$ is of rank ℓ. Define $\mathbf{P} = \mathbf{I} - \mathbf{A}^t(\mathbf{AA}^t)^{-1}\mathbf{A}$ as in Exercise 4.41 to be the matrix that projects any vector onto the nullspace of $\mathbf{A}$. Explain how Exercise 4.43 relates to checking the second-order necessary conditions for Problem P. How can you extend this to checking second-order sufficient conditions? Illustrate using Example 4.4.4.

[4.45] Consider the problem to maximize $3x_1x_2 + 2x_2x_3 + 12x_1x_3$ subject to $6x_1 + x_2 + 4x_3 = 6$. Using first- and second-order KKT optimality conditions, show that $\bar{\mathbf{x}} = (1/3, 2, 1/2)^t$ is a local maximum. Use Exercise 4.44 to check the second-order sufficient conditions.

[4.46] Consider the following problem:

$$\text{Minimize} \quad \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{Hx}$$
$$\text{subject to} \quad \mathbf{Ax} \leq \mathbf{b},$$

where $\mathbf{c}$ is an n-vector, $\mathbf{b}$ is an m-vector, $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{H}$ is an $n \times n$ symmetric matrix.

a. Write the second-order necessary optimality conditions of Theorem 4.4.3. Make all possible simplifications.
b. Is it necessarily true that every local minimum to the above problem is also a global minimum? Prove, or give a counter example.
c. Provide the first- and second-order necessary optimality conditions for the special case where $\mathbf{c} = \mathbf{0}$ and $\mathbf{H} = \mathbf{I}$. In this case the problem reduces to finding the point in a polyhedral set closest to the origin. (The above problem is referred to in the literature as a *least distance programming problem*.)

[4.47] Investigate the relationship between the optimal solutions and the KKT conditions for the following two problems, where $\lambda \geq \mathbf{0}$ is a given fixed vector.

P: Minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$, $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$.

P′: Minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$, $\lambda^t\mathbf{g}(\mathbf{x}) \leq 0$.

(Problem P′ has only one constraint and is referred to as a *surrogate relaxation* of Problem P.)

[4.48] Consider Problem P, to minimize $f(\mathbf{x})$ subject to $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$, where $f: R^n \to R$ and $h_i: R^n \to R$ for $i = 1,\ldots, \ell$ are all continuously differentiable functions. Let $\bar{\mathbf{x}}$ be a feasible solution and define an $\ell \times \ell$ Jacobian submatrix **J** as

$$\mathbf{J} = \begin{bmatrix} \frac{\partial h_1(\bar{\mathbf{x}})}{\partial x_1} & \frac{\partial h_1(\bar{\mathbf{x}})}{\partial x_\ell} \\ \vdots & \vdots \\ \frac{\partial h_\ell(\bar{\mathbf{x}})}{\partial x_1} & \frac{\partial h_\ell(\bar{\mathbf{x}})}{\partial x_\ell} \end{bmatrix}.$$

Assume that **J** is nonsingular, so, in particular, $\nabla h_i(\bar{\mathbf{x}})$, $i = 1,\ldots, \ell$, are linearly independent. Under these conditions, the *Implicit Function Theorem* (see Exercise 4.49) asserts that if we define $\mathbf{y} = (x_{\ell+1},\ldots,x_n)^t \in R^{n-\ell}$ with $\bar{\mathbf{y}} = (\bar{x}_{\ell+1},\ldots,\bar{x}_n)^t$, there exists a neighborhood of $\bar{\mathbf{y}}$ over which the (first) ℓ variables $x_1,\ldots,x_\ell$ can (implicitly) be solved for in terms of the y variables using the ℓ equality constraints. More precisely, there exists a neighborhood of $\bar{\mathbf{y}}$ and a set of functions $\psi_1(\mathbf{y}),\ldots,\psi_\ell(\mathbf{y})$ such that over this neighborhood, we have that $\psi_1(\mathbf{y}),\ldots,\psi_\ell(\mathbf{y})$ are continuously differentiable, $\psi_i(\bar{\mathbf{y}}) = \bar{x}_i$ for $i = 1,\ldots, \ell$, and $h_i[\psi_1(\mathbf{y}),\ldots,\psi_\ell(\mathbf{y}),\mathbf{y}] = 0$ for $i = 1,\ldots, \ell$.

Now suppose that $\bar{\mathbf{x}}$ is a local minimum and that the above assumptions hold true. Argue that $\bar{\mathbf{y}}$ must be a local minimum for the unconstrained function $F(\mathbf{y}) \equiv f[\psi_1(\mathbf{y}),\ldots,\psi_\ell(\mathbf{y}),\mathbf{y}]$: $R^{n-\ell} \to R$. Using the first-order necessary optimality conditions $\nabla F(\bar{\mathbf{y}}) = 0$ for unconstrained problems, derive the KKT necessary optimality conditions for Problem P. In particular, show that the Lagrangian multiplier vector $\bar{\mathbf{v}}$ in the dual feasibility condition $\nabla f(\bar{\mathbf{x}}) + [\nabla \mathbf{h}(\bar{\mathbf{x}})^t]\bar{\mathbf{v}}$, where $\nabla \mathbf{h}(\bar{\mathbf{x}})$ is the matrix whose rows are $\nabla h_1(\bar{\mathbf{x}})^t,\ldots,\nabla h_\ell(\bar{\mathbf{x}})^t$, is given uniquely by

$$\bar{\mathbf{v}} = -\mathbf{J}^{-1}\left[\frac{\partial f(\bar{\mathbf{x}})}{\partial x_1},\ldots,\frac{\partial f(\bar{\mathbf{x}})}{\partial x_\ell}\right]^t.$$

[4.49] Consider Problem P, to minimize $f(\mathbf{x})$ subject to $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$, where $\mathbf{x} \in R^n$ and where all objective and constraint functions are continuously differentiable. Suppose that $\bar{\mathbf{x}}$ is a local minimum for P and that the gradients $\nabla h_i(\bar{\mathbf{x}})$, $i = 1,\ldots, \ell$, are linearly independent. Use the implicit function theorem stated below to derive the KKT optimality conditions for P. Extend this to include inequality constraints $g_i(\mathbf{x}) \le 0$, $i = 1,..., m$ as well. (*Hint*: See Exercise 4.48.)

Implicit Function Theorem (see Taylor and Mann [1983], for example). Suppose that $\phi_i(\mathbf{x})$, $i = 1,..., p$ (representing the binding constraints at $\bar{\mathbf{x}}$), are continuously differentiable functions, and suppose that the gradients $\nabla\phi_i(\bar{\mathbf{x}})$, $i = 1,..., p$, are linearly independent, where $p < n$. Denote $\boldsymbol{\phi} \equiv (\phi_1,...,\phi_p)^t$: $R^n \to R^p$. Hence, $\boldsymbol{\phi}(\bar{\mathbf{x}}) = \mathbf{0}$ and we can partition $\mathbf{x}^t = (\mathbf{x}_B^t, \mathbf{x}_N^t)$, where $\mathbf{x}_B \in R^p$ and $\mathbf{x}_N \in R^{n-p}$ such that for the corresponding partition $[\nabla_B\boldsymbol{\phi}(\mathbf{x}), \nabla_N\boldsymbol{\phi}(\mathbf{x})]$ of the Jacobian $\nabla\boldsymbol{\phi}(\mathbf{x})$, the $p \times p$ submatrix $\nabla_B\boldsymbol{\phi}(\bar{\mathbf{x}})$ is nonsingular. Then the following holds true: There exists an open neighborhood $N_\varepsilon(\bar{\mathbf{x}}) \subseteq R^n$, $\varepsilon > 0$, an open neighborhood $N_{\varepsilon'}(\bar{\mathbf{x}}_N) \subseteq R^{n-p}$, $\varepsilon' > 0$, and a function $\boldsymbol{\psi}$: $R^{n-p} \to R^p$ that is continuously differentiable on $N_{\varepsilon'}(\bar{\mathbf{x}}_N)$ such that

(i) $\bar{\mathbf{x}}_B = \boldsymbol{\psi}(\bar{\mathbf{x}}_N)$.

(ii) For every $\mathbf{x}_N \in N_{\varepsilon'}(\bar{\mathbf{x}}_N)$, we have $\boldsymbol{\phi}[\boldsymbol{\psi}(\mathbf{x}_N), \mathbf{x}_N] = \mathbf{0}$.

(iii) The Jacobian $\nabla\boldsymbol{\phi}(\mathbf{x})$ has full row rank p for each $\mathbf{x} \in N_\varepsilon(\bar{\mathbf{x}})$.

(iv) For any $\mathbf{x}_N \in N_{\varepsilon'}(\bar{\mathbf{x}}_N)$, the Jacobian $\nabla\boldsymbol{\psi}(\mathbf{x}_N)$ is given by the (unique) solution to the linear equation system

$$\{\nabla_B\boldsymbol{\phi}[\nabla\boldsymbol{\psi}(\mathbf{x}_N), \mathbf{x}_N]\}\nabla\boldsymbol{\psi}(\mathbf{x}_N) = -\nabla_N\boldsymbol{\phi}[\boldsymbol{\psi}(\mathbf{x}_N), \mathbf{x}_N].$$

[4.50] A differentiable function ψ: $R^n \to R$ is said to be an *η-invex function* if there exists some (arbitrary) function η: $R^{2n} \to R^n$ such that for each $\mathbf{x}_1$, $\mathbf{x}_2 \in R^n$, $\psi(\mathbf{x}_2) \ge \psi(\mathbf{x}_1) + \nabla\psi(\mathbf{x}_1)^t \eta(\mathbf{x}_1, \mathbf{x}_2)$. Furthermore, ψ is said to be an *η-pseudoinvex function* if $\nabla\psi(\mathbf{x}_1)^t \eta(\mathbf{x}_1, \mathbf{x}_2) \ge 0$ implies that $\psi(\mathbf{x}_2) \ge \psi(\mathbf{x}_1)$. Similarly, ψ is said to be an *η-quasi-invex function* if $\psi(\mathbf{x}_2) \le \psi(\mathbf{x}_1)$ implies that $\nabla\psi(\mathbf{x}_1)^t \eta(\mathbf{x}_1, \mathbf{x}_2) \le 0$.

a. When invex is replaced by convex in the usual sense, what is $\eta(\mathbf{x}_1, \mathbf{x}_2)$ defined to be?

b. Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,..., m$ where f: $R^n \to R$ and g_i: $R^n \to R$ for $i = 1,..., m$ are all differentiable functions. Let $\bar{\mathbf{x}}$ be a KKT point. Show that $\bar{\mathbf{x}}$ is optimal if f and g_i for $i \in I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$ are all η-invex.

c. Repeat Part b if f is η-pseudoinvex and g_i, $i \in I$, are η-quasi-invex. (The reader is referred to Hanson [1981] and to Hanson and Mond [1982, 1987] for discussions on invex functions and their uses.)

Notes and References

In this chapter we begin by developing first- and second-order optimality conditions for unconstrained optimization problems in Section 4.1. These classical results can be found in most textbooks dealing with real analysis. For more details on this subject relating to higher-order necessary and sufficiency conditions, refer to Gue and Thomas [1968] and Hancock [1960]; and for information regarding the handling of equality constraints via the Lagrangian multiplier rule, refer to Bartle [1976] and Rudin [1964].

In Section 4.2 we treat the problem of minimizing a function in the presence of inequality constraints and develop the Fritz John [1948] necessary optimality conditions. A weaker form of these conditions, in which the nonnegativity of the multipliers was not asserted, was derived by Karush [1939]. Under a suitable constraint qualification, the Lagrangian multiplier associated with the objective function is positive, and the Fritz John conditions reduce to those of Kuhn and Tucker [1951], which were derived independently. Even though the latter conditions were originally derived by Karush [1939] using calculus of variations, this work had not received much attention, since it was not published. However, we refer to these conditions as KKT conditions, recognizing Karush, Kuhn, and Tucker. An excellent historical review of optimality conditions for nonlinear programming can be found in Kuhn [1976] and Lenstra et al. [1991]. Kyparisis [1985] presents a necessary and sufficient condition for the KKT Lagrangian multipliers to be unique. Gehner [1974] extends the FJ optimality conditions to the case of *semi-infinite nonlinear programming problems,* where there are an infinite number of parametrically described equality and inequality constraints. The reader may refer to the following references for further study of the Fritz John and KKT conditions: Abadie [1967b], Avriel [1967], Canon et al. [1966], Gould and Tolle [1972], Luenberger [1973], Mangasarian [1969a], and Zangwill [1969].

Mangasarian and Fromovitz [1967] generalized the Fritz John conditions for handling both equality and inequality constraints. Their approach used the implicit function theorem. In Section 4.3 we develop the Fritz John conditions for equality and inequality constraints by constructing a feasible arc, as in the work of Fiacco and McCormick [1968].

In Sections 4.2 and 4.3 we show that the KKT conditions are indeed sufficient for optimality under suitable convexity assumptions. This result was proved by Kuhn and Tucker [1951] if the functions f, g_i for $i \in I$ are convex, the functions h_i for all i are affine, and the set X is convex. This result was generalized later, so that weaker convexity assumptions are needed to guarantee optimality, as shown in Sections 4.2 and 4.3 (see Mangasarian [1969a]). The reader may also refer to Bhatt and Misra [1975], who relaxed the condition that h_i be affine, provided that the associated Lagrangian multiplier has the correct sign. Further generalizations using invex functions can be found in Hanson [1981] and Hanson and Mond [1982].

Other generalizations and extensions of the Fritz John and KKT conditions were developed by many authors. One such extension is to relax the condition that the set X is open. In this case we obtain necessary optimality conditions of the minimum principle type. For details on this type of optimality conditions, see Bazaraa and Goode [1972], Canon et al. [1970], and Mangasarian [1969a]. Another extension is to treat the problem in an infinite-dimensional setting. The interested reader may refer to Canon et al. [1970], Dubovitskii and Milyutin [1965], Gehner [1974], Guignard [1969], Halkin and Neustadt [1966], Hestenes [1966], Neustadt [1969], and Varaiya [1967].

In Section 4.4 we address second-order necessary and sufficient optimality conditions for constrained problems, developed initially by McCormick [1967]. Our second-order necessary optimality condition is stronger than that presented by McCormick [1967] (see Fletcher [1987] and Ben-Tal [1980]). For a discussion on checking these conditions based on eigenvalues computed over a projected tangential subspace, or based on bordered Hessian matrices, we refer the reader to Luenberger [1973a/1984]. See Exercise 4.44 for a related approach. For extensions and additional study of this topic, we refer the reader to Avriel [1976], Baccari and Trad [2004], Ben-Tal [1980], Ben-Tal and Zowe [1982], Fletcher [1983], Luenberger [1973a/1984], McCormick [1967], and Messerli and Polak [1969].

Chapter 5 Constraint Qualifications

In Chapter 4 we considered Problem P to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \le 0$, $i = 1,\dots, m$. We obtained the Karush–Kuhn–Tucker (KKT) necessary conditions for optimality by deriving the Fritz John conditions and then asserting that the multiplier associated with the objective function is positive at a local optimum when a constraint qualification is satisfied. In this chapter we develop the KKT conditions directly without first deriving the Fritz John conditions. This is done under various constraint qualifications for problems having inequality constraints and for problems having both inequality and equality constraints.

Following is an outline of the chapter.

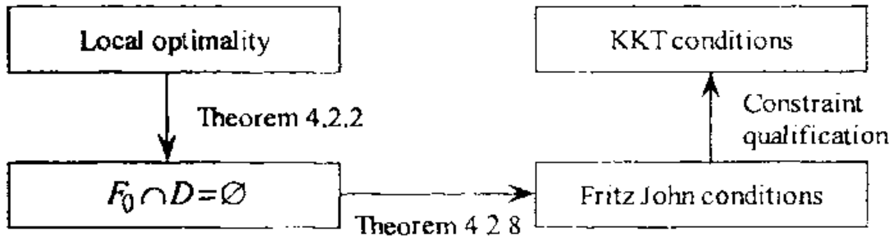
Section 5.1: Cone of Tangents We introduce the cone of tangents T and show that $F_0 \cap T = \varnothing$ is a necessary condition for local optimality. Using a constraint qualification, we derive the KKT conditions directly for problems having inequality constraints.

Section 5.2: Other Constraint Qualifications We introduce other cones contained in the cone of tangents. Making use of these cones, we present various constraint qualifications that validate the KKT conditions. Relationships among these constraint qualifications are explored.

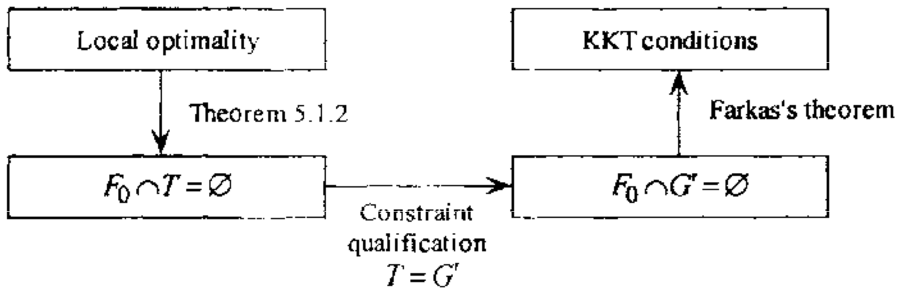
Section 5.3: Problems Having Inequality and Equality Constraints The results of Section 5.2 are extended to problems having equality and inequality constraints.

5.1 Cone of Tangents

In Section 4.2 we discussed the KKT necessary optimality conditions for problems having inequality constraints. In particular, we showed that local optimality implies that $F_0 \cap G_0 = \varnothing$, which in turn implies the Fritz John conditions. Under the linear independence constraint qualification, or, more generally, under the constraint qualification $G_0 \ne \varnothing$, we deduced that the Fritz John conditions can only be satisfied if the Lagrangian multiplier associated with the objective function is positive. This led to the KKT conditions. This process is summarized in the following flowchart.



In this section we derive the KKT conditions directly without first obtaining the Fritz John conditions. As shown in Theorem 5.1.2, a necessary condition for local optimality is that $F_0 \cap T = \varnothing$, where T is the cone of tangents introduced in Definition 5.1.1. Using the constraint qualification $T = G'$, where G' is as defined in Theorem 5.1.3 (see also Theorem 4.2.15), we get $F_0 \cap G' = \varnothing$. Using Farkas's theorem, this statement gives the KKT conditions. This process is summarized in the following flowchart.



5.1.1 Definition

Let S be a nonempty set in R^n, and let $\bar{\mathbf{x}} \in \operatorname{cl} S$. The *cone of tangents* of S at $\bar{\mathbf{x}}$, denoted by T, is the set of all directions $\mathbf{d}$ such that $\mathbf{d} = \lim_{k \to \infty} \lambda_k (\mathbf{x}_k - \bar{\mathbf{x}})$, where $\lambda_k > 0$, $\mathbf{x}_k \in S$ for each k, and $\mathbf{x}_k \to \bar{\mathbf{x}}$.

From the above definition, it is clear that $\mathbf{d}$ belongs to the cone of tangents if there is a feasible sequence $\{\mathbf{x}_k\}$ converging to $\bar{\mathbf{x}}$ such that the directions $\mathbf{x}_k - \bar{\mathbf{x}}$ converge to $\mathbf{d}$. Exercise 5.1 provides alternative equivalent descriptions for the cone of tangents T; and in Exercise 5.2, we ask the reader to show that the cone of tangents is indeed a closed cone. Figure 5.1 illustrates some examples of the cone of tangents, where the origin is translated to $\bar{\mathbf{x}}$ for convenience.

Theorem 5.1.2 shows that for a problem of the form to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$, $F_0 \cap T = \varnothing$ is, indeed, a necessary condition for optimality. Later we specify S to be the set $\{\mathbf{x} \in X : g_i(\mathbf{x}) \leq 0 \text{ for } i = 1,\ldots, m\}$.

5.1.2 Theorem

Let S be a nonempty set in R^n, and let $\bar{\mathbf{x}} \in S$. Furthermore, suppose that f: $R^n \to R$ is differentiable at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ locally solves the problem to minimize $f(\mathbf{x})$

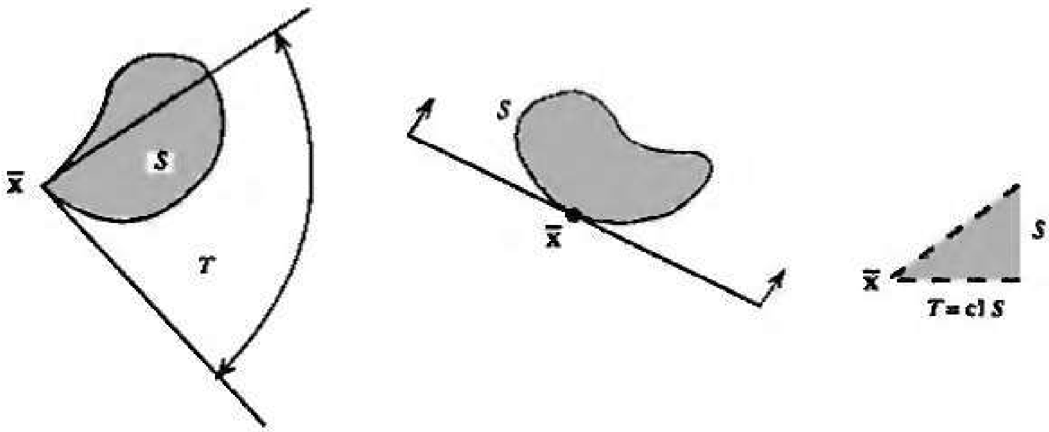


Figure 5.1 Cone of tangents.

subject to $\mathbf{x} \in S$, $F_0 \cap T = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\overline{\mathbf{x}})^t \mathbf{d} < 0\}$ and T is the cone of tangents of S at $\overline{\mathbf{x}}$.

Proof

Let $\mathbf{d} \in T$, that is, $\mathbf{d} = \lim_{k\to\infty} \lambda_k (\mathbf{x}_k - \overline{\mathbf{x}})$, where $\lambda_k > 0$, $\mathbf{x}_k \in S$ for each k, and $\mathbf{x}_k \to \overline{\mathbf{x}}$. By the differentiability of f at $\overline{\mathbf{x}}$, we get

$$f(\mathbf{x}_k) - f(\overline{\mathbf{x}}) = \nabla f(\overline{\mathbf{x}})^t (\mathbf{x}_k - \overline{\mathbf{x}}) + \|\mathbf{x}_k - \overline{\mathbf{x}}\| \alpha(\overline{\mathbf{x}}; \mathbf{x}_k - \overline{\mathbf{x}}), \tag{5.1}$$

where $\alpha(\overline{\mathbf{x}}; \mathbf{x}_k - \overline{\mathbf{x}}) \to 0$ as $\mathbf{x}_k \to \overline{\mathbf{x}}$. Noting the local optimality of $\overline{\mathbf{x}}$, we have for k large enough that $f(\mathbf{x}_k) \ge f(\overline{\mathbf{x}})$; so from (5.1) we get

$$\nabla f(\overline{\mathbf{x}})^t (\mathbf{x}_k - \overline{\mathbf{x}}) + \|\mathbf{x}_k - \overline{\mathbf{x}}\| \alpha(\overline{\mathbf{x}}; \mathbf{x}_k - \overline{\mathbf{x}}) \ge 0.$$

Multiplying by $\lambda_k > 0$ and taking the limit as $k \to \infty$, the above inequality implies that $\nabla f(\overline{\mathbf{x}})^t \mathbf{d} \ge 0$. Hence, we have shown that $\mathbf{d} \in T$ implies that $\nabla f(\overline{\mathbf{x}})^t \mathbf{d} \ge 0$ and therefore that $F_0 \cap T = \varnothing$. This completes the proof.

It is worth noting that the condition $F_0 \cap T = \varnothing$ does not necessarily imply that $\overline{\mathbf{x}}$ is a local minimum. Indeed, this condition will hold true whenever $F_0 = \varnothing$, for example, which we know is not sufficient for local optimality. However, if there exists an ε-neighborhood $N_\varepsilon(\overline{\mathbf{x}})$ about $\overline{\mathbf{x}}$ such that $N_\varepsilon(\overline{\mathbf{x}}) \cap S$ is convex and f is pseudoconvex over $N_\varepsilon(\overline{\mathbf{x}}) \cap S$, $F_0 \cap T = \varnothing$ is sufficient to claim that $\overline{\mathbf{x}}$ is a local minimum (see Exercise 5.3).

Abadie Constraint Qualification

In Theorem 5.1.3, we derive the KKT conditions under the constraint qualification $T = G'$, which is credited to Abadie.

5.1.3 Theorem (Karush–Kuhn–Tucker Necessary Conditions)

Let X be a nonempty set in R^n, and let f: $R^n \to R$ and g_i: $R^n \to R$ for $i = 1,\ldots, m$. Consider the problem: Minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \leq 0$ for $i = 1,\ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible solution, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{\mathbf{x}}$. Furthermore, suppose that the constraint qualification $T = G'$ holds true, where T is the cone of tangents of the feasible region at $\bar{\mathbf{x}}$ and $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \leq 0 \text{ for } i \in I\}$. If $\bar{\mathbf{x}}$ is a local optimal solution, there exist nonnegative scalars u_i for $i \in I$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

Proof

By Theorem 5.1.2 we have that $F_0 \cap T = \varnothing$, where $F_0 = \{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0\}$. By assumption, $T = G'$, so that $F_0 \cap G' = \varnothing$. In other words, the following system has no solution:

$$\nabla f(\bar{\mathbf{x}})^t \mathbf{d} < 0, \qquad \nabla g_i(\bar{\mathbf{x}})^t \mathrm{d} \leq 0 \qquad \text{for } i \in I.$$

Hence, by Theorem 2.4.5 (Farkas's theorem), the result follows (see also Theorem 4.2.15).

The reader may verify that in Example 4.2.10, the constraint qualification $T = G'$ does not hold true at $\bar{\mathbf{x}} = (1,0)^t$. Note that the Abadie constraint qualification $T = G'$ could be stated equivalently as $T \supseteq G'$, since $T \subseteq G'$ is always true (see Exercise 5.4). Note that openness of the set X and continuity of g_i at $\bar{\mathbf{x}}$ for $i \notin I$ were not assumed explicitly in Theorem 5.1.3. However, without these assumptions, it is unlikely that the constraint qualification $T \supseteq G'$ would hold true (see Exercise 5.5).

Linearly Constrained Problems

Lemma 5.1.4 shows that if the constraints are linear, the Abadie constraint qualification is satisfied automatically. This also implies that the KKT conditions are always necessary for problems having linear constraints, whether the objective function is linear or nonlinear. As an alternative proof that does not employ the cone of tangents, note that if $\bar{\mathbf{x}}$ is a local minimum, $F_0 \cap D = \varnothing$. Now, by Lemma 4.2.4, if the constraints are linear, $D = G'_0 \equiv \{\mathbf{d} \neq \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \leq 0, \text{ for each } i \in I\}$. Hence, $F_0 \cap D = \varnothing \Leftrightarrow F_0 \cap G'_0 = \varnothing$, which holds true if and only if $\bar{\mathbf{x}}$ is a KKT point by Theorem 4.2.15.

5.1.4 Lemma

Let $\mathbf{A}$ be an $m \times n$ matrix, let $\mathbf{b}$ be an m-vector, and let $S = \{\mathbf{x} : \mathbf{Ax} \le \mathbf{b}\}$. Suppose that $\bar{\mathbf{x}} \in S$ is such that $\mathbf{A}_1\bar{\mathbf{x}} = \mathbf{b}_1$ and $\mathbf{A}_2\bar{\mathbf{x}} < \mathbf{b}_2$, where $\mathbf{A}^t = (\mathbf{A}_1^t, \mathbf{A}_2^t)$ and $\mathbf{b}^t = (\mathbf{b}_1^t, \mathbf{b}_2^t)$. Then $T = G'$, where T is the cone of tangents of S at $\bar{\mathbf{x}}$ and $G' = \{\mathrm{d} : \mathbf{A}_1\mathrm{d} \le \mathbf{0}\}$.

Proof

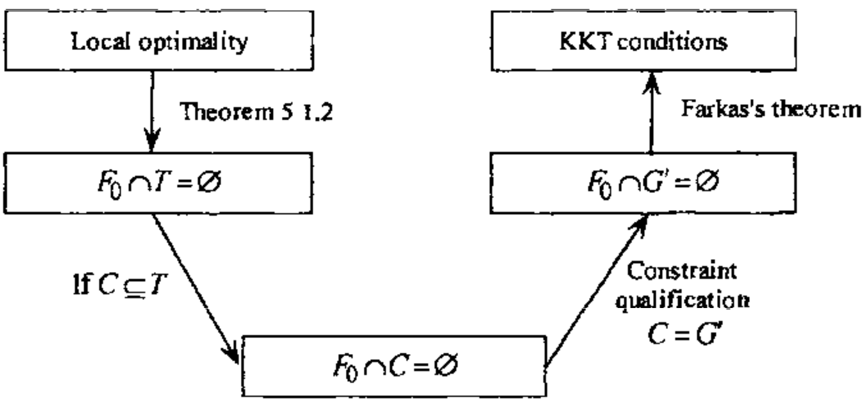
If $\mathbf{A}_1$ is vacuous, then $\mathbf{G}' = R^n$. Furthermore, $\bar{\mathbf{x}} \in \text{int } S$ and hence $T = R^n$. Thus, $G' = T$. Now, suppose that $\mathbf{A}_1$ is not vacuous. Let $\mathbf{d} \in T$; that is, $\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathrm{x}_k - \bar{\mathrm{x}})$, where $\mathrm{x}_k \in S$ and $\lambda_k > 0$ for each k. Then

$$\mathrm{A}_1(\mathrm{x}_k - \bar{\mathrm{x}}) \le \mathbf{b}_1 - \mathbf{b}_1 = \mathbf{0}. \tag{5.2}$$

Multiplying (5.2) by $\lambda_k > 0$ and taking the limit as $k \to \infty$, it follows that $\mathbf{A}_1\mathbf{d} \le \mathbf{0}$. Thus, $\mathbf{d} \in G'$ and $T \subseteq G'$. Now let $\mathbf{d} \in G'$; that is, $\mathbf{A}_1\mathbf{d} \le \mathbf{0}$. We need to show that $\mathbf{d} \in T$. Since $\mathrm{A}_2\bar{\mathrm{x}} < \mathbf{b}_2$, there is a $\delta > 0$ such that $\mathbf{A}_2(\bar{\mathbf{x}} + \lambda\mathbf{d}) < \mathbf{b}_2$ for all $\lambda \in (0, \delta)$. Furthermore, since $\mathbf{A}_1\bar{\mathrm{x}} = \mathbf{b}_1$ and $\mathbf{A}_1\mathbf{d} \le \mathbf{0}$, then $\mathrm{A}_1(\bar{\mathbf{x}} + \lambda\mathbf{d}) \le \mathbf{b}_1$ for all $\lambda > 0$. Therefore, $\bar{\mathrm{x}} + \lambda\mathbf{d} \in S$ for each $\lambda \in (0, \delta)$. This shows automatically that $\mathbf{d} \in T$. Therefore, $T = G'$, and the proof is complete.

5.2 Other Constraint Qualifications

The KKT conditions have been developed by many authors under various constraint qualifications. In this section we present some of the more important constraint qualifications. In Section 5.1 we learned that local optimality implies that $F_0 \cap T = \varnothing$ and that the KKT conditions follow under the constraint qualification $T = G'$. If we define a cone $C \subseteq T$, $F_0 \cap T = \varnothing$ also implies that $F_0 \cap C = \varnothing$. Therefore, any constraint qualification of the form $C = G'$ will lead to the KKT conditions. In fact, since $C \subseteq T \subseteq G'$, the constraint qualification $C = G'$ implies that $T = G'$ and is therefore more restrictive than Abadie's constraint qualification. This process is illustrated in the following flowchart:



We present below several such cones whose closures are contained in T. Here the feasible region S is given by $\{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0, i = 1,..., m\}$. The vector $\bar{\mathbf{x}}$ is a feasible point, and $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$.

Cone of Feasible Directions of S at $\bar{\mathbf{x}}$

This cone was introduced in Definition 4.2.1. The cone of feasible directions, denoted by D, is the set of all nonzero vectors $\mathbf{d}$ such that $\bar{\mathbf{x}} + \lambda\mathbf{d} \in S$ for $\lambda \in (0, \delta)$ for some $\delta > 0$.

Cone of Attainable Directions of S at $\bar{\mathbf{x}}$

A nonzero vector $\mathbf{d}$ belongs to the cone of attainable directions, denoted by A, if there exist a $\delta > 0$ and an $\boldsymbol{\alpha}: R \to R^n$ such that $\boldsymbol{\alpha}(\lambda) \in S$ for $\lambda \in (0, \delta)$, $\boldsymbol{\alpha}(0) = \bar{\mathbf{x}}$, and

$$\lim_{\lambda \to 0^+} \frac{\boldsymbol{\alpha}(\lambda) - \boldsymbol{\alpha}(0)}{\lambda} = \mathbf{d}.$$

In other words, $\mathbf{d}$ belongs to the cone of attainable directions if there is a feasible arc starting from $\bar{\mathbf{x}}$ that is tangential to $\mathbf{d}$.

Cone of Interior Directions of S at $\bar{\mathbf{x}}$

This cone, denoted by G_0, was introduced in Section 4.2 and is defined as $G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} < 0 \text{ for } i \in I\}$. Note that if X is open and each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, then $\mathbf{d} \in G_0$ implies that $\bar{\mathbf{x}} + \lambda\mathbf{d}$ belongs to the interior of the feasible region for $\lambda > 0$ and sufficiently small.

Lemma 5.2.1 shows that all the above cones and their closures are contained in T.

5.2.1 Lemma

Let X be a nonempty set in R^n, and let $f: R^n \to R$ and $g_i: R^n \to R$ for $i = 1,..., m$. Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,..., m$ and $\mathbf{x} \in X$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that each g_i for $i \in I$ is differentiable at $\bar{\mathbf{x}}$, and let $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0 \text{ for } i \in I\}$. Then

$$\text{cl}\, D \subseteq \text{cl}\, A \subseteq T \subseteq G',$$

where D, A, and T are, respectively, the cone of feasible directions, the cone of attainable directions, and the cone of tangents of the feasible region at $\bar{\mathbf{x}}$. Furthermore, if X is open and each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, then $G_0 \subseteq D$, so that

$$\text{cl}\, G_0 \subseteq \text{cl}\, D \subseteq \text{cl}\, A \subseteq T \subseteq G',$$

where G_0 is the cone of interior directions of the feasible region at $\bar{\mathbf{x}}$.

Proof

It can easily be verified that $D \subseteq A \subseteq T \subseteq G'$ and that, since T is closed (see Exercise 5.2), $\text{cl}\, D \subseteq \text{cl}\, A \subseteq T \subseteq G'$. Now note that $G_0 \subseteq D$ by Lemma 4.2.4. Hence, the second part of the lemma follows.

To illustrate how each of the containments considered above can be strict, consider the following example. In Figure 4.9, note that $G_0 = \varnothing = \text{cl}\, G_0$, since there are no interior directions, whereas $D = \text{cl}\, D = G'$ is defined by the feasible direction along the edge incident at $\bar{\mathbf{x}}$. In regard to the cone of interior directions G_0, note that whereas any $\mathbf{d} \in G_0$ is a direction leading to interior feasible solutions, it is not true that any feasible direction that leads to interior points belongs to G_0. For example, consider Example 4.3.5, illustrated in Figure 4.12, with the equalities replaced by "less than or equal to" inequalities. The set $G_0 = \varnothing$ at $\bar{\mathbf{x}} = (1,0)^t$, whereas $\mathbf{d} = (-1,0)^t$ leads to interior feasible points.

To show that cl D can be a strict subset of cl A, consider the region defined by $x_1 - x_2^2 \le 0$ and $-x_1 + x_2^2 \le 0$. The set of feasible points lies on the parabola $x_1 = x_2^2$. At $\bar{\mathbf{x}} = (0,0)^t$, for example, $D = \varnothing = \text{cl}\, D$, whereas cl $A = \{\mathbf{d} : \mathbf{d} = \lambda(0,1)^t \text{ or } \mathbf{d} = \lambda(0,-1)^t,\ \lambda \ge 0\} = G'$.

The possibility that cl $A \ne T$ is a little more subtle. Suppose that the feasible region S is itself the sequence $\{(1/k,0)^t,\ k = 1, 2, ...\}$ formed by the intersection of suitable constraints (written as suitable inequalities). For example, we might have $S = \{(x_1, x_2) : x_2 = h(x_1),\ x_2 = 0,\ 0 \le x_1 \le 1$, where $h(x_1) = x_1^3 \sin(\pi/x_1)$ if $x_1 \ne 0$ and $h(x_1) = 0$ if $x_1 = 0\}$. Then $A = \varnothing = \text{cl}\, A$, since there are no feasible arcs. However, by definition, $T = \{\mathbf{d} : \mathbf{d} = \lambda(1,0)^t,\ \lambda \ge 0\}$, and it is readily verified that $T = G'$.

Finally, Figure 4.7 illustrates an instance where T is a strict subset of G'. Here $T = \{\mathbf{d} : \mathbf{d} = \lambda(-1,0)^t,\ \lambda \ge 0\}$, while $G' = \{\mathbf{d} : \mathbf{d} = \lambda(-1,0)^t, \text{ or } \mathbf{d} = \lambda(1,0)^t,\ \lambda \ge 0\}$.

We now present some constraint qualifications that validate the KKT conditions and discuss their interrelationships.

Slater's Constraint Qualification

The set X is open, each g_i for $i \in I$ is pseudoconvex at $\bar{\mathbf{x}}$, for each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and there is an $\mathbf{x} \in X$ such that $g_i(\mathbf{x}) < 0$ for all $i \in I$.

Linear Independence Constraint Qualification

The set X is open, each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ are linearly independent.

Cottle's Constraint Qualification

The set X is open and each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, and $\text{cl}\, G_0 = G'$.

Zangwill's Constraint Qualification

$$\text{cl}\, D = G'.$$

Kuhn–Tucker's Constraint Qualification

$$\text{cl}\, A = G'.$$

Validity of the Constraint Qualifications and Their Interrelationships

In Theorem 5.1.3 we showed that the KKT necessary optimality conditions are necessarily true under Abadie's constraint qualification $T = G'$. We demonstrate below that all the constraint qualifications discussed above imply that of Abadie and hence, each validate the KKT necessary conditions. From Lemma 5.2.1 it is clear that Cottle's constraint qualification implies that of Zangwill, which implies that of Kuhn and Tucker, which in turn implies Abadie's qualification. We now show that the first two constraint qualifications imply that of Cottle.

First, suppose that Slater's constraint qualification holds true. Then there is an $\mathbf{x} \in X$ such that $g_i(\mathbf{x}) < 0$ for $i \in I$. Since $g_i(\mathbf{x}) < 0$ and $g_i(\bar{\mathbf{x}}) = 0$, then, by the pseudoconvexity of g_i at $\bar{\mathbf{x}}$, it follows that $\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) < 0$. Thus, d $= \mathbf{x} - \bar{\mathbf{x}}$ belongs to G_0. Therefore, $G_0 \neq \varnothing$ and the reader can verify that $\text{cl}\, G_0 = G'$ and hence that Cottle's constraint qualification holds true. Now, suppose that the linear independence constraint qualification is satisfied. Then $\sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}$ has no nonzero solution. By Theorem 2.4.9 it follows that there exists a vector d such that $\nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0$ for all $i \in I$. Thus, $G_0 \neq \varnothing$, and Cottle's constraint qualification holds true. The relationships among the foregoing constraint qualifications are illustrated in Figure 5.2.

In discussing Lemma 5.2.1, we gave various examples in which for each consecutive pair in the string of containments $\text{cl}\, G_0 \subseteq \text{cl}\, D \subseteq \text{cl}\, A \subseteq T \subseteq G'$, the containment was strict and the larger set was equal to G'. Hence, these examples also illustrate that the implications of Figure 5.2 in regard to these sets are one-way implications. Figure 5.2 accordingly illustrates for each constraint qualification an instance where it holds true, whereas the preceding constraint qualification that makes a more restrictive assumption does not hold true.

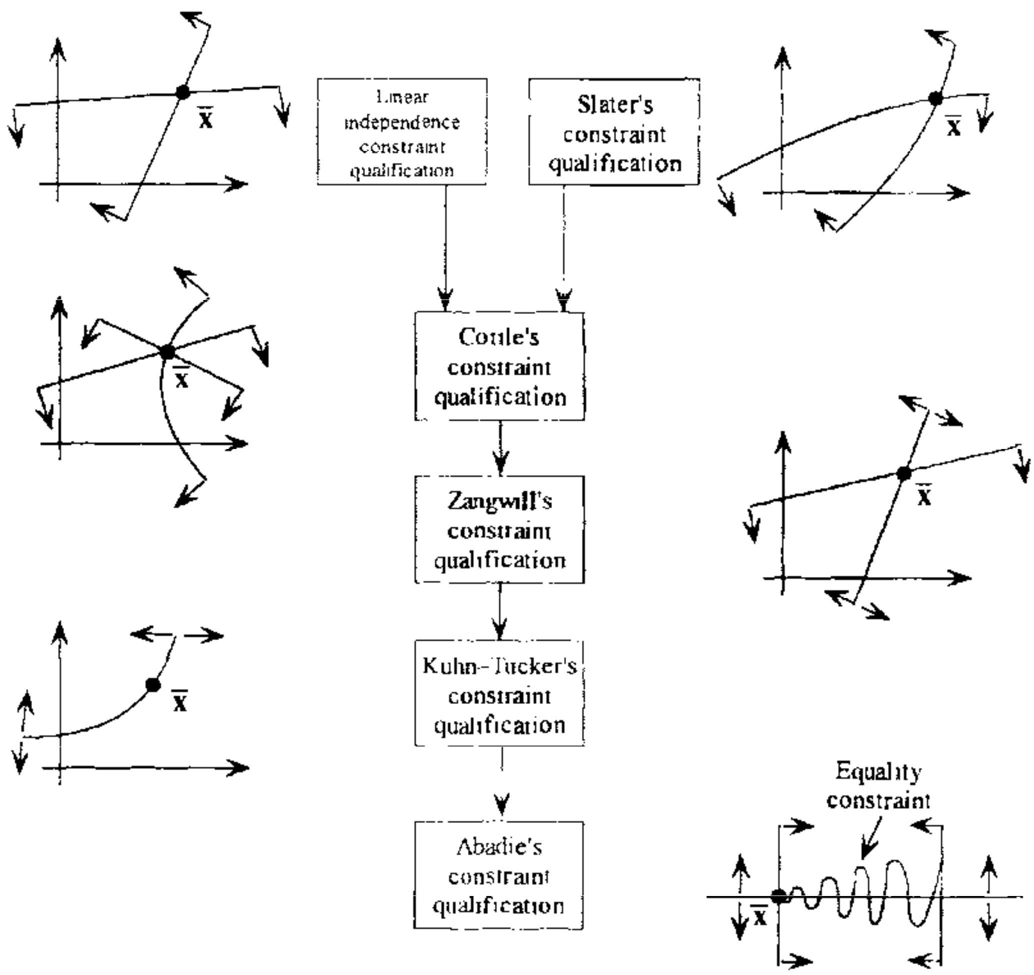


Figure 5.2 Relationships among various constraint qualifications for inequality-constrained problems.

Needless to say, if a local minimum $\bar{x}$ is not a KKT point, as in Example 4.2.10 and illustrated in Figure 4.7, for instance, no constraint qualification can possibly hold true.

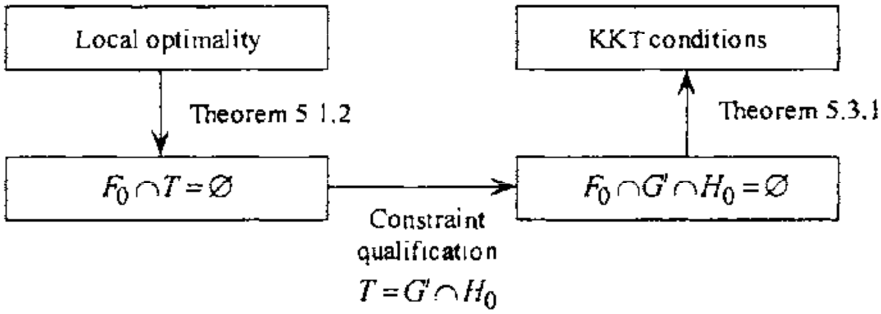
Finally, we remark that Cottle's constraint qualification is equivalent to requiring that $G_0 \neq \varnothing$ (see Exercise 5.6). Moreover, we have seen that Slater's constraint qualification and the linear independence constraint qualification both imply Cottle's constraint qualification. Hence, whenever these constraint qualifications hold true at a local minimum $\bar{x}$, then $\bar{x}$ is a Fritz John point, with the Lagrangian multiplier u_0 associated with the objective function necessarily positive. In contrast, we might have Zangwill's, or Kuhn–Tucker's, or Abadie's constraint qualifications holding true at a local minimum $\bar{x}$, while u_0 might possibly be zero in some solution to the Fritz John conditions. However, since these are valid constraint qualifications, in such a case we must also have $u_0 > 0$ in some solution to the Fritz John conditions.

5.3 Problems Having Inequality and Equality Constraints

In this section we consider problems having both inequality and equality constraints. In particular, consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \qquad \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 \qquad \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. \end{aligned}$$

By Theorem 5.1.2 a necessary optimality condition is $F_0 \cap T = \varnothing$ at a local minimum $\bar{\mathbf{x}}$. By imposing the constraint qualification $T = G' \cap H_0$, where $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\ldots,\ell\}$, this implies that $F_0 \cap G' \cap H_0 = \varnothing$. By using Farkas's theorem [or see Equation (4.16)], the KKT conditions follow. Theorem 5.3.1 reiterates this. The process is summarized in the following flowchart:



5.3.1 Theorem (Karush–Kuhn–Tucker Conditions)

Let f: $R^n \to R$, g_i: $R^n \to R$ for $i = 1,\ldots, m$, and h_i: $R^n \to R$ for $i = 1,\ldots, \ell$, and let X be a nonempty set in R^n. Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \qquad \text{for } i = 1,\ldots,m \\ & h_i(\mathbf{x}) = 0 \qquad \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. \end{aligned}$$

Let $\bar{\mathbf{x}}$ locally solve the problem, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that f, g_i for $i \in I$, and h_i for $i = 1,\ldots, \ell$ are differentiable at $\bar{\mathbf{x}}$. Suppose that the constraint qualification $T = G' \cap H_0$ holds true, where T is the cone of tangents of the feasible region at $\bar{\mathbf{x}}$, and

$$\begin{aligned} G' &= \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for } i \in I\}, \\ H_0 &= \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\ldots,\ell\}. \end{aligned}$$

Then $\bar{\mathbf{x}}$ is a KKT point; that is, there exist scalars $u_i \ge 0$ for $i \in I$ and v_i for $i = 1,\ldots, \ell$ such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

Proof

Since $\bar{\mathbf{x}}$ solves the problem locally, by Theorem 5.1.2 we have that $F_0 \cap T = \varnothing$. By the constraint qualification, we have $F_0 \cap G' \cap H_0 = \varnothing$; that is, the system $\mathbf{Ad} \le \mathbf{0}$ and $\mathbf{c}^t\mathbf{d} > 0$ has no solution, where the rows of $\mathbf{A}$ are given by $\nabla g_i(\bar{\mathbf{x}})^t$ for $i \in I$, $\nabla h_i(\bar{\mathbf{x}})^t$ and $-\nabla h_i(\bar{\mathbf{x}})^t$ for $i = 1,\ldots,\ell$, and $\mathbf{c} = -\nabla f(\bar{\mathbf{x}})$. By Theorem 2.4.5, the system $\mathbf{A}^t\mathbf{y} = \mathbf{c}$ and $\mathbf{y} \ge \mathbf{0}$ has a solution. This implies that there exist nonnegative scalars u_i for $i \in I$ and α_i, β_i for $i = 1,\ldots,\ell$, such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \alpha_i \nabla h_i(\bar{\mathbf{x}}) - \sum_{i=1}^{\ell} \beta_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}.$$

Letting $v_i = \alpha_i - \beta_i$ for each i, the result follows.

We now present several constraint qualifications that validate the KKT conditions. These qualifications use several cones that were defined earlier in the chapter. By replacing each equality constraint by two equivalent inequalities, the role played by G' in the preceding section is now played by the cone $G' \cap H_0$. The reader may note that Zangwill's constraint qualification is omitted here, since the cone of feasible directions is usually equal to the zero vector in the presence of nonlinear equality constraints.

Slater's Constraint Qualification

The set X is open, each g_i for $i \in I$ is pseudoconvex at $\bar{\mathbf{x}}$, each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, each h_i for $i = 1,\ldots,\ell$ is quasiconvex, quasiconcave, and continuously differentiable at $\bar{\mathbf{x}}$, and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots,\ell$ are linearly independent. Furthermore, there exists an $\mathbf{x} \in X$ such that $g_i(\mathbf{x}) < 0$ for all $i \in I$ and $h_i(\mathbf{x}) = 0$ for all $i = 1,\ldots,\ell$.

Linear Independence Constraint Qualification

The set X is open, each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots,\ell$, are linearly independent, and each h_i for $i = 1,\ldots,\ell$ is continuously differentiable at $\bar{\mathbf{x}}$.

Cottle's Constraint Qualification

The set X is open, each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$, each h_i for $i = 1,\ldots, \ell$ is continuously differentiable at $\bar{\mathbf{x}}$, and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots, \ell$ are linearly independent. Furthermore, $\text{cl}\ (G_0 \cap H_0) = G' \cap H_0$. [This is equivalent to the *Mangasarian–Fromovitz constraint qualification*, which requires $\nabla h_i(\bar{\mathbf{x}})$, $i = 1,\ldots, \ell$, to be linearly independent and that $G_0 \cap H_0 \neq \varnothing$; see Exercise 5.7.]

Kuhn–Tucker's Constraint Qualification

$$\text{cl}\ A = G' \cap H_0.$$

Abadie's Constraint Qualification

$$T = G' \cap H_0.$$

Validity of the Constraint Qualifications and Their Interrelationships

In Theorem 5.3.1 we showed that the KKT conditions hold true if Abadie's constraint qualification $T = G' \cap H_0$ is satisfied. We demonstrate below that all the constraint qualifications given above imply that of Abadie, and hence, each validates the KKT necessary conditions.

As in Lemma 5.2.1, the reader can easily verify that $\text{cl}\ A \subseteq T \subseteq G' \cap H_0$. Now, suppose that X is open, g_i for each $i \notin I$ is continuous at $\bar{\mathbf{x}}$, h_i for each $i = 1,\ldots, \ell$ is continuously differentiable, and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots, \ell$ are linearly independent. From the proof of Theorem 4.3.1 it follows that $G_0 \cap H_0 \subseteq A$. Thus, $\text{cl}\ (G_0 \cap H_0) \subseteq \text{cl}\ A \subseteq T \subseteq G' \cap H_0$. In particular, Cottle's constraint qualification implies that of Kuhn and Tucker, which in turn implies Abadie's constraint qualification.

We now demonstrate that Slater's constraint qualification and the linear independence constraint qualification imply that of Cottle. Suppose that Slater's qualification is satisfied, so that $g_i(\mathbf{x}) < 0$ for $i \in I$ and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$ for some $\mathbf{x} \in X$. By the pseudoconvexity of g_i at $\bar{\mathbf{x}}$, we get that $\nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) < 0$ for $i \in I$.

Also, since $h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}}) = 0$, the quasiconvexity and quasiconcavity of h_i at $\bar{\mathbf{x}}$ imply that $\nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) = 0$. Letting $\mathbf{d} = \mathbf{x} - \bar{\mathbf{x}}$, it follows that $\mathbf{d} \in G_0 \cap H_0$. Thus, $G_0 \cap H_0 \neq \varnothing$, and the reader can verify that $\text{cl}(G_0 \cap H_0) = G' \cap H_0$. Therefore, Cottle's constraint qualification holds true.

Finally, we show that the linear independence constraint qualification implies that of Cottle. By contradiction, suppose that $G_0 \cap H_0 = \varnothing$. Then, using a separation theorem as in the proof of Theorem 4.3.2, it follows that there exists a nonzero vector $(\mathbf{u}_I, \mathbf{v})$ such that $\sum_{i\in I} u_i \nabla g_i(\overline{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{\mathbf{x}}) = \mathbf{0}$, where $\mathbf{u}_I \geq \mathbf{0}$ is the vector whose *i*th component is u_i for $i \in I$. This contradicts the linear independence assumption. Thus, Cottle's constraint qualification holds true.

In Figure 5.3 we summarize the implications of the constraint qualifications discussed above (see also Figure 5.2). As mentioned earlier, these implications, together with Theorem 5.3.1, validate the KKT conditions.

Second-Order Constraint Qualifications for Inequality- and Equality-Constrained Problems

In Chapter 4 we developed second-order necessary KKT optimality conditions. In particular, we observed in Theorem 4.4.3 that if $\overline{\mathbf{x}}$ is a local minimum and if all problem-defining functions are twice differentiable, with the gradients $\nabla g_i(\overline{\mathbf{x}})$ for $i \in I$ and $\nabla h_i(\overline{\mathbf{x}})$ for $i = 1,..., \ell$ of the binding constraints being linearly independent, then $\overline{\mathbf{x}}$ is a KKT point and, additionally, $\mathbf{d}^t \nabla^2 L(\overline{\mathbf{x}}) \mathbf{d} \geq 0$ must hold true for all $\mathbf{d} \in C$ as defined therein. Hence, the linear independence condition affords a *second-order constraint qualification*, which implies that in addition to $\overline{\mathbf{x}}$ being a KKT point, a second-order type of condition must also hold true.

Alternatively, we can stipulate the following second-order constraint qualification, which is in the spirit of Abadie's constraint qualification. Suppose that all problem-defining functions are twice differentiable and that $\overline{\mathbf{x}}$ is a local

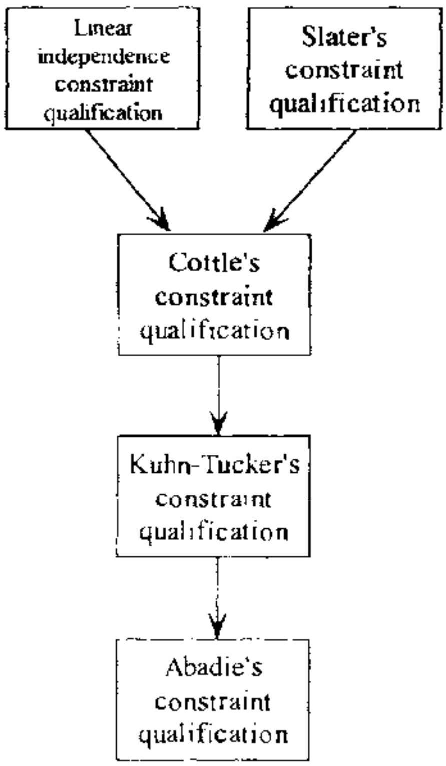


Figure 5.3 Relationships among constraint qualifications for problems having inequality and equality constraints.

minimum at which Abadie's constraint qualification $T = G' \cap H_0$ holds true. Hence, we know from Theorem 5.3.1 that $\bar{x}$ is a KKT point. Denote by $\bar{u}$ and $\bar{v}$ the associated set of Lagrangian multipliers corresponding to the inequality and equality constraints, respectively, and let $I = \{i : g_i(\bar{x}) = 0\}$ represent the binding inequality constraints. Now, as in Theorem 4.4.3, define $C = \{\mathbf{d} \neq \mathbf{0}: \nabla g_i(\bar{x})^t \mathbf{d} = 0$ for $i \in I^+$, $\nabla g_i(\bar{x})^t \mathbf{d} \leq \mathbf{0}$ for $i \in I^0$, and $\nabla h_i(\bar{x})^t \mathbf{d} = 0$ for $i = 1,\ldots, \ell\}$, where $I^+ = \{i \in I : \bar{u}_i > 0\}$ and $I^0 = I - I^+$.

Accordingly, let T' denote the cone of tangents at $\bar{x}$ when the inequality constraints with indices $i \in I^+$ are also treated as equalities. Then the stated second-order constraint qualification asserts that if $T' = C \cup \{\mathbf{0}\}$, we must have $\mathbf{d}^t\nabla^2 L(\bar{x})\mathbf{d} \geq 0$ for each $\mathbf{d} \in C$. We ask the reader to show that this assertion is valid in Exercise 5.9, using a proof similar to that of Theorem 4.4.3. Note that, in general, $T' \subseteq C \cup \{\mathbf{0}\}$, in the same manner as $T \subseteq G' \cap H_0$. However, as evident from the proof of Theorem 4.4.3, under the linear independence constraint qualification, any $\mathbf{d} \in C$ also corresponds to a limiting direction based on a feasible arc and hence based on a sequence of points. Therefore, $C \cup \{\mathbf{0}\} \subseteq T'$. This shows that the linear independence constraint qualification implies that $T' = C \cup \{\mathbf{0}\}$. We ask the reader to construct the precise details of this argument in Exercise 5.9. In a similar manner, we can state another second-order constraint qualification in the spirit of Kuhn–Tucker's (cone of attainable directions) constraint qualification. This is addressed in Exercise 5.10, where we ask the reader to justify it and to show that this is also implied by the linear independence constraint qualification.

Exercises

[5.1] Prove that the cone of tangents defined in Definition 5.1.1 can be characterized equivalently in either of the following ways:

a. $T = \{\mathbf{d}$: there exists a sequence $\{\lambda_k\} \to 0^+$ and a function α: $R \to R^n$, where $\alpha(\lambda) \to 0$ as $\lambda \to 0$, such that $\mathbf{x}_k = \bar{\mathbf{x}} + \lambda_k \mathbf{d} + \lambda_k \alpha(\lambda_k) \in S$ for each $k\}$.

b. $T = \{\mathbf{d} : \mathbf{d} = \lambda \lim_{k\to\infty} \dfrac{\mathbf{x}_k - \bar{\mathbf{x}}}{\|\mathbf{x}_k - \bar{\mathbf{x}}\|}$, where $\lambda \geq 0$, $\{\mathbf{x}_k\} \to \bar{\mathbf{x}}$, and where $\mathbf{x}_k \in S$ and $\mathbf{x}_k \neq \bar{\mathbf{x}}$, for each $k\}$.

[5.2] Prove that the cone of tangents is a closed cone. [*Hint*: First show that $T = \bigcap_{N \in \mathcal{N}} \text{cl}\, K(S \cap N, \bar{\mathbf{x}})$, where $K(S \cap N, \bar{\mathbf{x}}) = \{\lambda(\mathbf{x} - \bar{\mathbf{x}}) : \mathbf{x} \in S \cap N, \lambda > 0\}$ and $\mathcal{N}$ is the class of all open neighborhoods about $\bar{\mathbf{x}}$.]

[5.3] For a nonlinear optimization problem, let $\bar{x}$ be a feasible solution, let F be the set of improving directions, let $F_0 = \{\mathrm{d} : \nabla f(\bar{x})^t \mathrm{d} < 0\}$, and let T be the cone of tangents at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is a local minimum, is $F \cap T = \varnothing$? Is $F \cap T = \varnothing$ sufficient to claim that $\bar{\mathbf{x}}$ is a local minimum? Give examples to justify your answers. Show that if there exists an ε-neighborhood about $\bar{\mathbf{x}}$ over which f is pseudoconvex and the feasible region is a convex set, then $F_0 \cap T = \varnothing$ implies that $\bar{\mathbf{x}}$ is a local minimum, so these conditions also guarantee that $\bar{\mathbf{x}}$ is a local minimum whenever $F \cap T = \varnothing$.

[5.4] Let $S = \{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0 \text{ for } i = 1,\dots, m\}$. Let $\bar{\mathbf{x}} \in S$, and let $I = \{i : g_i(\mathbf{x}) = 0\}$. Show that $T \subseteq G'$, where T is the cone of tangents of S at $\bar{\mathbf{x}}$, and $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for } i \in I\}$.

[5.5] Consider the problem to maximize $5x - x^2$ subject to $g_1(x) \le 0$, where $g_1(x) = x$.

a. Verify graphically that $\bar{x} = 0$ is the optimal solution.
b. Verify that each of the constraint qualifications discussed in Section 5.2 holds true at $\bar{x} = 0$.
c. Verify that the KKT necessary conditions hold true at $\bar{x} = 0$.

Now suppose that the constraint $g_2(x) \le 0$ is added to the above problem, where

$$g_2(x) = \begin{cases} -1 - x & \text{if } x \ge 0 \\ 1 - x & \text{if } x < 0. \end{cases}$$

Note that $\bar{x} = 0$ is still the optimal solution and that g_2 is discontinuous and nonbinding at $\bar{x}$. Check whether the constraint qualifications discussed in Section 5.2 and the KKT conditions hold true at $\bar{x}$. (This exercise illustrates the need of the continuity assumption of the nonbinding constraints.)

[5.6] Let **A** be an $m \times n$ matrix, and consider the cones $G_0 = \{\mathbf{d} : \mathbf{A}\mathbf{d} < \mathbf{0}\}$ and $G' = \{\mathbf{d} : \mathbf{A}\mathbf{d} \le \mathbf{0}\}$. Prove that:

a. G_0 is an open convex cone.
b. G' is a closed convex cone.
c. $G_0 = \text{int } G'$.
d. $\text{cl } G_0 = G'$ if and only if $G_0 \neq \varnothing$.

[5.7] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\dots, m$, $h_i(\mathbf{x}) = 0$ for $i = 1,\dots, \ell$, and $\mathbf{x} \in X$, where X is open and where all problem-defining functions are differentiable. Let $\bar{\mathbf{x}}$ be a feasible solution. The *Mangasarian–Fromovitz constraint qualification* requires that $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\dots, \ell$ are linearly independent and that $G_0 \cap H_0 \neq \varnothing$, where $G_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} < 0 \text{ for } i \in I\}$, $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$, and $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,\dots, \ell\}$. Show that $G_0 \cap H_0 \neq \varnothing$ if and only if $\text{cl}(G_0 \cap H_0) = G' \cap H_0$, where $G' = \{\mathbf{d} :$

$\nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0$ for $i \in I\}$ and hence that this constraint qualification is equivalent to that of Cottle's.

[5.8] Let S be a subset of R^n, and let $\bar{\mathbf{x}} \in \text{int } S$. Show that the cone of tangents of S at $\bar{\mathbf{x}}$ is R^n.

[5.9] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,..., m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,..., \ell$, where all problem-defining functions are twice differentiable. Let $\bar{\mathbf{x}}$ be a local minimum, and suppose that the cone of tangents $T = G' \cap H_0$, where $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for } i \in I\}$, $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$, and $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,..., \ell\}$. Hence, by Theorem 5.3.1, $\bar{\mathbf{x}}$ is a KKT point. Let $\bar{u}_i$, $i = 1,..., m$, and $\bar{v}_i$, $i = 1,..., \ell$, be the associated Lagrangian multipliers in the KKT solution with respect to the inequality and the equality constraints, and define $I^+ = \{i : \bar{u}_i > 0\}$, and let $I^0 = I - I^+$. Now define T' as the cone of tangents at $\bar{\mathbf{x}}$ with respect to the region $\{\mathbf{x} : g_i(\mathbf{x}) = 0 \text{ for } i \in I^+, g_i(\mathbf{x}) \le 0 \text{ for } i \in I^0, h_i(\mathbf{x}) = 0 \text{ for } i = 1,..., \ell\}$, and denote $C = \{\mathbf{d} \ne \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i \in I^+, \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for } i \in I^0, \text{ and } \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,..., \ell\}$. Show that if $T' = C \cup \{\mathbf{0}\}$, the second-order necessary condition $\mathbf{d}^t \nabla^2 L(\bar{\mathbf{x}}) \mathbf{d} \ge 0$ holds true for all $\mathbf{d} \in C$, where $L(\mathbf{x}) = f(\mathbf{x}) + \sum_{i \in I} \bar{u}_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} \bar{v}_i h_i(\mathbf{x})$. Also, show that the linear independence constraint qualification implies that $T' = C \cup \{\mathbf{0}\}$. (*Hint*: Examine the proof of Theorem 4.4.3.)

[5.10] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,..., m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,..., \ell$, where all problem-defining functions are twice differentiable. Let $\bar{\mathbf{x}}$ be a local minimum that is also a KKT point. Define $\bar{C} = \{\mathbf{d} \ne \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i \in I, \text{ and } \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0 \text{ for } i = 1,..., \ell\}$, where $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. The *second-order cone of attainable directions constraint qualification* is said to hold true at $\bar{\mathbf{x}}$ if every $\mathbf{d} \in \bar{C}$ is tangential to a twice differentiable arc incident at $\bar{\mathbf{x}}$; that is, for every $\mathbf{d} \in \bar{C}$, there exists a twice differentiable function $\boldsymbol{\alpha} : [0, \varepsilon] \to R^n$ for some $\varepsilon > 0$, such that for each $0 \le \lambda \le \varepsilon$, $\boldsymbol{\alpha}(0) = \bar{\mathbf{x}}$, $g_i[\boldsymbol{\alpha}(\lambda)] = 0$ for $i \in I$, $h_i[\boldsymbol{\alpha}(\lambda)] = 0$ for $i = 1,..., \ell$, and $\lim_{\lambda \to 0^+} [\boldsymbol{\alpha}(\lambda) - \boldsymbol{\alpha}(0)]/\lambda = \theta \mathbf{d}$ for some $\theta > 0$. Assuming that this condition holds true, show that $\mathbf{d}^t \nabla^2 L(\bar{\mathbf{x}}) \mathbf{d} \ge 0$ for all $\mathbf{d} \in \bar{C}$, where $L(\bar{\mathbf{x}})$ is defined by (4.25). Also, show that this second-order constraint qualification is implied by the linear independence constraint qualification.

[5.11] Find the cone of tangents for each of the following sets at the point $\bar{\mathbf{x}} = (0,0)^t$:

a. $S = \{(x_1, x_2) : x_2 \ge -x_1^3\}$.

b. $S = \{(x_1, x_2) : x_1 \text{ is integer, } x_2 = 0\}$.

c. $S = \{(x_1, x_2) : x_1 \text{ is rational, } x_2 = 0\}$.

[5.12] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$. Let $\bar{\mathbf{x}}$ be feasible, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Let $(\bar{z}, \bar{\mathbf{d}})$ be an optimal solution to the following linear program:

$$\begin{aligned} \text{Minimize} \quad & z \\ \text{subject to} \quad & \nabla f(\bar{\mathbf{x}})^t \mathbf{d} - z \le 0 \\ & \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} - z \le 0 && \text{for } i \in I \\ & -1 \le d_j \le 1 && \text{for } j = 1, \ldots, n. \end{aligned}$$

a. Show that the Fritz John conditions hold true if $\bar{z} = 0$.
b. Show that if $\bar{z} = 0$, the KKT conditions hold true under Cottle's constraint qualification .

[5.13] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & -x_1 \\ \text{subject to} \quad & x_1^2 + x_2^2 \le 1 \\ & (x_1 - 1)^3 - x_2 \le 0. \end{aligned}$$

a. Show that the Kuhn–Tucker constraint qualification holds true at $\bar{\mathbf{x}} = (1,0)^t$.

b. Show that $\bar{\mathbf{x}} = (1,0)^t$ is a KKT point and that it is the global optimal solution.

[5.14] For each of the following sets, find the cone of feasible directions and the cone of attainable directions at $\bar{\mathbf{x}} = (0,0)^t$:

a. $S = \{(x_1, x_2) : -1 \le x_1 \le 1, \;\; x_2 \ge x_1^{1/3}, \;\; x_2 \ge x_1\}$.

b. $S = \{(x_1, x_2) : x_2 > x_1^2\}$.

c. $S = \{(x_1, x_2) : x_2 = -x_1^3\}$.

d. $S = S_1 \cup S_2$, where

$$S_1 = \{(x_1, x_2) : x_1 \ge 0, \;\; x_2 \ge x_1^2\} \qquad \text{and}$$
$$S_2 = \{(x_1, x_2) : x_1 \le 0, \;\; -2x_1 \le 3x_2 \le -x_1\}.$$

[5.15] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$. Let $\bar{\mathbf{x}}$ be a feasible point, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Suppose that X is open and each g_i for $i \notin I$ is continuous at $\bar{\mathbf{x}}$. Further, suppose that the set

$$\{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0 \text{ for } i \in J,\ \nabla g_i(\bar{\mathbf{x}})^t \mathrm{d} < 0 \text{ for } i \in I - J\}$$

is not empty, where $J = \{i \in I : g_i$ is pseudoconcave at $\bar{\mathbf{x}}\}$. Show that this condition is sufficient to validate the KKT conditions at $\bar{\mathbf{x}}$. (This is the *Arrow–Hurwicz–Uzawa constraint qualification.*)

[5.16] Let $f: R^n \to R$ be differentiable at $\bar{\mathbf{x}}$ with a nonzero gradient $\nabla f(\bar{\mathbf{x}})$. Let $S = \{\mathbf{x} : f(\mathbf{x}) \ge f(\bar{\mathbf{x}})\}$. Show that the cone of tangents and the cone of attainable directions of S at $\bar{\mathbf{x}}$ are both given by $\{\mathbf{d} : \nabla f(\bar{\mathbf{x}})^t \mathbf{d} \ge 0\}$. Does this result hold true if $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$? Prove or give a counterexample.

[5.17] Consider the feasible region $S = \{\mathbf{x} \in X : g_1(\mathbf{x}) \le 0\}$, where $g_1(\mathbf{x}) = x_1^2 + x_2^2 - 1$ and X is the collection of all convex combinations of the four points $(-1,0)^t$, $(0,1)^t$, $(1,0)^t$, and $(0,-1)^t$.

a. Find the cone of tangents T of S at $\bar{\mathbf{x}} = (1,0)^t$.

b. Check whether $T \supseteq G'$, where $G' = \{\mathbf{d} : \nabla g_1(\bar{\mathbf{x}})^t \mathbf{d} \le 0\}$.

c. Replace the set X by four inequality constraints. Repeat parts a and b, where $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathbf{d} \le 0$ for $i \in I\}$ and I is the new set of binding constraints at $\bar{\mathbf{x}} = (1,0)^t$.

[5.18] Let $S = \{\mathbf{x} \in S : g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell\}$. Let $\bar{\mathbf{x}} \in S$, and let $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Show that $T \subseteq G' \cap H_0$, where T is the cone of tangents of S at $\bar{\mathbf{x}}$, $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t \mathrm{d} \le 0$ for $i \in I\}$, and $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t \mathbf{d} = 0$ for $i = 1,\ldots, \ell\}$.

[5.19] Consider Abadie's constraint qualification $T = G' \cap H_0$ for the case of inequality and equality constraints. Using the Kuhn–Tucker example of Figure 4.7, for instance, demonstrate by considering differentiable objective functions that it is valid and more general to instead require $F_0 \cap T = F_0 \cap G' \cap H_0$ to guarantee that if $\bar{\mathbf{x}}$ is a local minimum, it is a KKT point. (Typically, "constraint qualifications" address only the behavior of the constraints and neglect the objective function.) Investigate the KKT conditions and Abadie's constraint qualification for the problem to minimize $\{f(\mathbf{x}) : g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell\}$ versus those for the equivalent problem to minimize $\{z : f(\mathbf{x}) \le z,\ g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$, and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell\}$.

[5.20] Consider the constraints $\mathbf{Cd} \le \mathbf{0}$ and $\mathbf{d}^t\mathbf{d} \le 1$. Let $\bar{\mathbf{d}}$ be a feasible solution such that $\bar{\mathbf{d}}^t\bar{\mathbf{d}} = 1$, $\mathbf{C}_1\bar{\mathbf{d}} = \mathbf{0}$, and $\mathbf{C}_2\bar{\mathbf{d}} < \mathbf{0}$, where $\mathbf{C}^t = (\mathbf{C}_1^t, \mathbf{C}_2^t)$. Show that the constraint qualification $T = G_1 = \{\mathbf{d} : \mathbf{C}_1\mathbf{d} \le \mathbf{0}, \mathbf{d}^t\bar{\mathbf{d}} \le 0\}$ holds true, where T is the cone of tangents of the constraint set at $\bar{\mathbf{d}}$.

[5.21] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$ and $h_i(\mathbf{x}) = 0$ for $i = 1,\ldots, \ell$, where all problem-defining functions are twice differentiable. Let $\bar{\mathbf{x}}$ be a local minimum that also happens to be a KKT point with associated Lagrangian multipliers vectors $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ corresponding to the inequality and equality constraints, respectively. Define $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$, $I^+ = \{i : \bar{u}_i > 0\}$, $I^0 = I - I^+$, $\bar{G}_0 = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} < 0$ for $i \in I^0$, $\nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$ for $i \in I^+\}$, and $H_0 = \{\mathbf{d} : \nabla h_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$ for $i = 1,\ldots, \ell\}$. The *strict Mangasarian–Fromovitz constraint qualification* (SMFCQ) is said to hold true at $\bar{\mathbf{x}}$ if $\nabla g_i(\bar{\mathbf{x}})$ for $i \in I^+$ and $\nabla h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots, \ell$ are linearly independent and $\bar{G}_0 \cap H_0 \ne \varnothing$. Show that the SMFCQ condition holds true at $\bar{\mathbf{x}}$ if and only if the KKT Lagrangian multiplier vector $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ is unique. Moreover, show that if the SMFCQ condition holds true at $\bar{\mathbf{x}}$, then $\mathbf{d}^t\nabla^2 L(\bar{\mathbf{x}})\mathbf{d} \ge 0$ for all $\mathbf{d} \in C = \{\mathbf{d} \ne \mathbf{0} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$ for $i \in I^+$, $\nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0$ for $i \in I^0$, and $\nabla h_i(\bar{\mathbf{x}})^t\mathbf{d} = 0$ for $i = 1,\ldots, \ell\}$, where $L(\mathbf{x}) = f(\mathbf{x}) + \sum_{i \in I} \bar{u}_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} \bar{v}_i h_i(\mathbf{x})$. (This result is due to Kyparisis [1985] and Ben-Tal [1980].)

[5.22] Consider the feasible region defined by $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$, where g_i: $R^n \to R$, $i = 1,\ldots, m$, are differentiable functions. Let $\bar{\mathbf{x}}$ be a feasible solution, denote by Ξ the set of differentiable objective functions f: $R^n \to R$ for which $\bar{\mathbf{x}}$ is a local minimum, and let $D\Xi = \{\mathbf{y} : \mathbf{y} = \nabla f(\bar{\mathbf{x}})$ for some $f \in \Xi\}$. Define the set $G' = \{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0$ for $i \in I\}$, where $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$, and let T be the cone of tangents at $\bar{\mathbf{x}}$. Furthermore, for any set S, let S_* denote its reverse polar cone defined as $\{\mathbf{y} : \mathbf{y}^t\mathbf{x} \ge 0$ for all $\mathbf{x} \in S\}$.

a. Show that $D\Xi = T_*$.
b. Show that the KKT conditions hold true for all $f \in \Xi$ if and only if $T_* = G'_*$. (*Hint:* The statement in Part b occurs if and only if $D\Xi \subseteq G'_*$. Now, use Part a along with the fact that $T_* \supseteq G'_*$ since $T \subseteq G'$. This result is due to Gould and Tolle [1971, 1972].)

Notes and References

In this chapter we provide an alternative derivation of the KKT conditions for problems having inequality constraints and problems having both equality and inequality constraints. This is done directly by imposing a suitable constraint qualification, as opposed to first developing the Fritz John conditions and then the KKT conditions.

The KKT optimality conditions were originally developed by imposing the constraint qualification that for every direction vector d in the cone G', there is a feasible arc whose tangent at $\bar{x}$ points along d. Since then, many authors have developed the KKT conditions under different constraint qualifications. For a thorough study of this subject, refer to the works of Abadie [1967b], Arrow et al. [1961], Canon et al. [1966], Cottle [1963a], Evans [1970], Evans and Gould [1970], Guignard [1969], Mangasarian [1969a], Mangasarian and Fromovitz [1967], and Zangwill [1969]. For a comparison and further study of these constraint qualifications, see the survey articles of Bazaraa et al. [1972], Gould and Tolle [1972], and Peterson [1973].

Gould and Tolle [1971] showed that the constraint qualification of Guignard [1969] is the weakest possible in the sense that it is both necessary and sufficient for the validation of the KKT conditions (see Exercise 5.22 for a precise statement). For further discussion on constraint qualifications that validate second-order necessary optimality conditions, refer to Ben-Tal [1980], Ben-Tal and Zowe [1982], Fletcher [1987], Kyparisis [1985], and McCormick [1967]. Also, for an application of KKT conditions under various constraint qualifications in conducting *sensitivity analyses* in nonlinear programming, see Fiacco [1983].

Chapter 6 Lagrangian Duality and Saddle Point Optimality Conditions

Given a nonlinear programming problem, there is another nonlinear programming problem closely associated with it. The former is called the *primal problem*, and the latter is called the *Lagrangian dual problem*. Under certain convexity assumptions and suitable constraint qualifications, the primal and dual problems have equal optimal objective values and, hence, it is possible to solve the primal problem indirectly by solving the dual problem.

Several properties of the dual problem are developed in this chapter. They are used to provide general solution strategies for solving the primal and dual problems. As a by-product of one of the duality theorems, we obtain saddle point necessary optimality conditions without any differentiability assumptions.

Following is an outline of the chapter.

Section 6.1: Lagrangian Dual Problem We introduce the Lagrangian dual problem, give its geometric interpretation, and illustrate it by several numerical examples.

Section 6.2: Duality Theorems and Saddle Point Optimality Conditions We prove the weak and strong duality theorems. The latter shows that the primal and dual objective values are equal under suitable convexity assumptions. We also develop the saddle point optimality conditions along with necessary and sufficient conditions for the absence of a duality gap, and interpret this in terms of a suitable perturbation function.

Section 6.3: Properties of the Dual Function We study several important properties of the dual function, such as concavity, differentiability, and subdifferentiability. We then give necessary and sufficient characterizations of ascent and steepest ascent directions.

Section 6.4: Formulating and Solving the Dual Problem Several procedures for solving the dual problem are discussed. In particular, we describe briefly gradient and subgradient-based methods, and present a tangential approximation cutting plane algorithm.

Section 6.5: Getting the Primal Solution We show that the points generated during the course of solving the dual problem yield optimal solutions to perturbations of the primal problem. For convex programs, we show how to obtain primal feasible solutions that are near-optimal.

Section 6.6: Linear and Quadratic Programs We give Lagrangian dual formulations for linear and quadratic programming, relating them to other standard duality formulations

6.1 Lagrangian Dual Problem

Consider the following nonlinear programming Problem P, which we call the *primal problem.*

Primal Problem P

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1,\dots,m \\ & h_i(\mathbf{x}) = 0 \quad \text{for } i = 1,\dots,\ell \\ & \mathbf{x} \in X. \end{aligned}$$

Several problems, closely related to the above primal problem, have been proposed in the literature and are called *dual problems.* Among the various duality formulations, the Lagrangian duality formulation has perhaps attracted the most attention. It has led to several algorithms for solving large-scale linear problems as well as convex and nonconvex nonlinear problems. It has also proved useful in discrete optimization where all or some of the variables are further restricted to be integers. The *Lagrangian dual problem D* is stated below.

Lagrangian Dual Problem D

$$\begin{aligned} \text{Maximize} \quad & \theta(\mathbf{u}, \mathbf{v}) \\ \text{subject to} \quad & \mathbf{u} \ge 0, \end{aligned}$$

where $\theta(\mathbf{u}, \mathbf{v}) = \inf\{f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} v_i h_i(\mathbf{x}) : \mathbf{x} \in X\}$.

Note that the *Lagrangian dual function* θ may assume the value of $-\infty$ for some vectors (u, v). The optimization problem that evaluates $\theta(\mathbf{u}, \mathbf{v})$ is sometimes referred to as the *Lagrangian dual subproblem.* In this problem the constraints $g_i(\mathbf{x}) \le 0$ and $h_i(\mathbf{x}) = 0$ have been incorporated in the objective function using the *Lagrangian multipliers* or *dual variables* u_i and v_i, respectively. This process of accommodating the constraints within the objective function using dual or Lagrangian multipliers is referred to as *dualization.* Also note that the multiplier u_i associated with the inequality constraint $g_i(\mathbf{x}) \le 0$ is nonnegative, whereas the multiplier v_i associated with the equality constraint $h_i(\mathbf{x}) = 0$ is unrestricted in sign.

Since the dual problem consists of maximizing the infimum (greatest lower bound) of the function $f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} v_i h_i(\mathbf{x})$, it is sometimes referred to as the *max–min dual problem.* We remark here that strictly speaking, we should write D as $\sup\{\theta(\mathrm{u},\mathrm{v}) : \mathrm{u} \geq 0\}$, rather than $\max\{\theta(\mathrm{u},\mathrm{v}) : \mathrm{u} \geq 0\}$, since the maximum may not exist (see Example 6.2.8). However, we shall specifically identify such cases wherever necessary.

The primal and Lagrangian dual problems can be written in the following form using vector notation, where f: $R^n \to R$, g: $R^n \to R^m$ is a vector function whose ith component is g_i, and h: $R^n \to R^\ell$ is a vector function whose ith component is h_i. For the sake of convenience, we shall use this form throughout the remainder of this chapter.

Primal Problem P

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in X. \end{aligned}$$

Lagrangian Dual Problem D

$$\begin{aligned} \text{Maximize} \quad & \theta(\mathbf{u},\mathbf{v}) \\ \text{subject to} \quad & \mathbf{u} \geq \mathbf{0}, \end{aligned}$$

where $\theta(\mathbf{u},\mathbf{v}) = \inf\{f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}$.

Given a nonlinear programming problem, several Lagrangian dual problems can be devised, depending on which constraints are handled as $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and which constraints are treated by the set X. This choice can affect both the optimal value of D (as in nonconvex situations) and the effort expended in evaluating and updating the dual function θ during the course of solving the dual problem. Hence, an appropriate selection of the set X must be made, depending on the structure of the problem and the purpose for solving D (see the Notes and References section).

Geometric Interpretation of the Dual Problem

We now discuss briefly the geometric interpretation of the dual problem. For the sake of simplicity, we consider only one inequality constraint and assume that no equality constraints exist. Then the primal problem is to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g(\mathbf{x}) \leq 0$.

In the (y, z) plane, the set $\{(y,z) : y = g(\mathbf{x}),\ z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in X\}$ is denoted by G in Figure 6.1. Thus, G is the image of X under the (g, f) map. The primal problem asks us to find a point in G with $y \leq 0$ that has a minimum ordinate. Obviously, this point is $(\bar{y},\ \bar{z})$ in Figure 6.1.

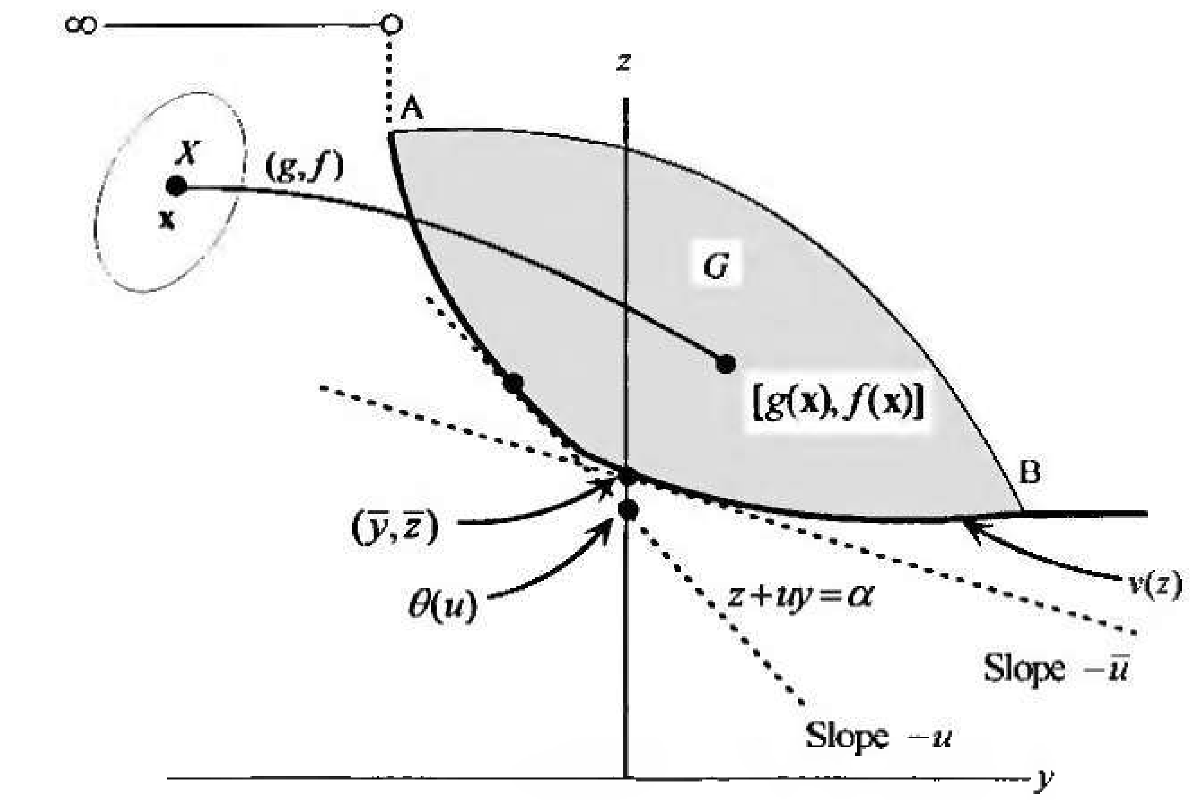


Figure 6.1 Geometric interpretation of Lagrangian duality.

Now suppose that $u \geq 0$ is given. To determine $\theta(u)$, we need to minimize $f(\mathbf{x})+ug(\mathbf{x})$ over all $\mathbf{x} \in X$. Letting $y = g(\mathbf{x})$ and $z = f(\mathbf{x})$ for $\mathbf{x} \in X$, we want to minimize $z + uy$ over points in G. Note that $z + uy = \alpha$ is an equation of a straight line with slope $-u$ and intercept α on the z-axis. To minimize $z + uy$ over G, we need to move the line $z + uy = \alpha$ parallel to itself as far down (along its negative gradient) as possible while it remains in contact with G. In other words, we move this line parallel to itself until it supports G from below, that is, the set G lies above the line and touches it. Then the intercept on the z-axis gives $\theta(u)$, as shown in Figure 6.1. The dual problem is therefore equivalent to finding the slope of the supporting hyperplane such that its intercept on the z-axis is maximal. In Figure 6.1, such a hyperplane has slope $-\bar{u}$ and supports the set G at the point $(\bar{y}, \bar{z})$. Thus, the optimal dual solution is $\bar{u}$, and the optimal dual objective value is $\bar{z}$. Furthermore, the optimal primal and dual objectives are equal in this case.

There is a related interesting interpretation that provides an important conceptual tool in this context. For the problem under consideration, define the function

$$v(y) = \min\{f(\mathbf{x}) : g(\mathbf{x}) \leq y, \ \mathbf{x} \in X\}.$$

The function v is called a *perturbation function* since it is the optimal value function of a problem obtained from the original problem by perturbing the right-hand side of the inequality constraint $g(\mathbf{x}) \leq 0$ to y from the value of zero. Note that $v(y)$ is a nonincreasing function of y since, as y increases, the feasible region of the perturbed problem enlarges (or stays the same). For the present case, this function is illustrated in Figure 6.1. Observe that v corresponds here to the lower envelope of G between points A and B because this envelope is itself monotone decreasing. Moreover, v remains constant at the value at point B for values of y higher than that at B and becomes ∞ for points to the left of A because of infeasibility. In particular, if v is differentiable at the origin, we observe that $v'(0) = -\bar{u}$. Hence, the marginal rate of change in objective function value with an increase in the right-hand side of the constraint from its present value of zero is given by $-\bar{u}$, the negative of the Lagrangian multiplier value at optimality. If v is convex but is not differentiable at the origin, then $-\bar{u}$ is evidently a subgradient of v at $y = 0$. In either case we know that $v(y) \ge v(0) - \bar{u}y$ for all $y \in R$. As we shall see later, v can be nondifferentiable and/or nonconvex, but the condition $v(y) \ge v(0) - \bar{u}y$ holds true for all $y \in R$ if and only if $\bar{u}$ is a KKT Lagrangian multiplier corresponding to an optimal solution $\bar{\mathbf{x}}$ such that it solves the dual problem with equal primal and dual objective values. As seen above, this happens to be the case in Figure 6.1.

6.1.1 Example

Consider the following primal problem:

$$\begin{aligned} \text{Minimize} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & -x_1 - x_2 + 4 \le 0 \\ & x_1, x_2 \ge 0. \end{aligned}$$

Note that the optimal solution occurs at the point $(x_1, x_2) = (2, 2)$, whose objective value is equal to 8.

Letting $g(x) = -x_1 - x_2 + 4$ and $X = \{(x_1, x_2) : x_1, x_2 \ge 0\}$, the dual function is given by

$$\begin{aligned} \theta(u) &= \inf\{x_1^2 + x_2^2 + u(-x_1 - x_2 + 4) : x_1, x_2 \ge 0\} \\ &= \inf\{x_1^2 - ux_1 : x_1 \ge 0\} + \inf\{x_2^2 - ux_2 : x_2 \ge 0\} + 4u. \end{aligned}$$

Note that the above infima are achieved at $x_1 = x_2 = u/2$ if $u \ge 0$ and at $x_1 = x_2 = 0$ if $u < 0$. Hence,

$$\theta(u) = \begin{cases} -\dfrac{1}{2}u^2 + 4u & \text{for } u \ge 0 \\ 4u & \text{for } u < 0. \end{cases}$$

Note that θ is a concave function, and its maximum over $u \ge 0$ occurs at $\bar{u} = 4$. Figure 6.2 illustrates the situation. Note also that the optimal primal and dual objective values are both equal to 8.

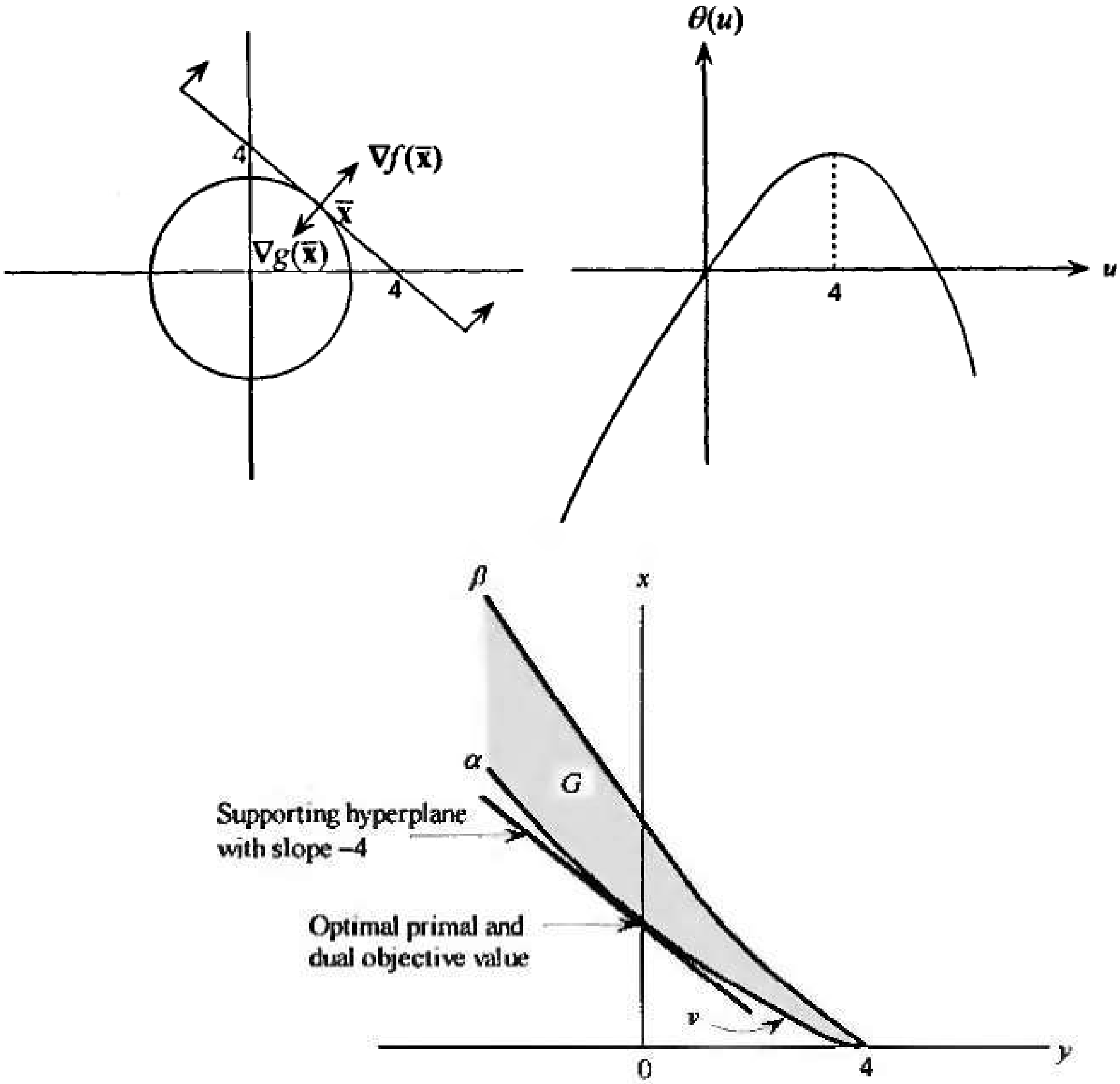


Figure 6.2 Geometric interpretation of Example 6.1.1.

Now let us consider the problem in the (y, z)-plane, where $y = g(x)$ and $z = f(x)$. We are interested in finding G, the image of $X = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\}$, under the (g, f) map. We do this by deriving explicit expressions for the lower and upper envelopes of G, denoted, respectively, by α and β.

Given y, note that $\alpha(y)$ and $\beta(y)$ are the optimal objective values of the following problems P_1 and P_2, respectively.

Problem P_1		Problem P_2	
Minimize	$x_1^2 + x_2^2$	Maximize	$x_1^2 + x_2^2$
subject to	$-x_1 - x_2 + 4 = y$	subject to	$-x_1 - x_2 + 4 = y$
	$x_1, x_2 \geq 0.$		$x_1, x_2 \geq 0.$

The reader can verify that $\alpha(y) = (4-y)^2/2$ and $\beta(y) = (4-y)^2$ for $y \le 4$. The set G is illustrated in Figure 6.2. Note that $\mathbf{x} \in X$ implies that $x_1,\ x_2 \ge 0$, so that $-x_1 - x_2 + 4 \le 4$. Thus, every point $\mathbf{x} \in X$ corresponds to $y \le 4$.

Note that the optimal dual solution is $\bar{u} = 4$, which is the negative of the slope of the supporting hyperplane shown in Figure 6.2. The optimal dual objective value is $\alpha(0) = 8$ and is equal to the optimal primal objective value.

Again, in Figure 6.2, the perturbation function $v(y)$ for $y \in R$ corresponds to the lower envelope $\alpha(y)$ for $y \le 4$, and $v(y)$ remains constant at the value 0 for $y \ge 4$. The slope $v'(0)$ equals -4, the negative of the optimal Lagrange multiplier value. Moreover, we have $v(y) \ge v(0) - 4y$ for all $y \in R$. As we shall see in the next section, this is a necessary and sufficient condition for the primal and dual objective values to match at optimality.

6.2 Duality Theorems and Saddle Point Optimality Conditions

In this section we investigate the relationships between the primal and dual problems and develop saddle point optimality conditions for the primal problem.

Theorem 6.2.1, referred to as the *weak duality theorem*, shows that the objective value of any feasible solution to the dual problem yields a lower bound on the objective value of any feasible solution to the primal problem. Several important results follow as corollaries.

6.2.1 Theorem (Weak Duality Theorem)

Let $\mathbf{x}$ be a feasible solution to Problem P; that is, $\mathbf{x} \in X$, $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$, and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. Also, let $(\mathbf{u}, \mathbf{v})$ be a feasible solution to Problem D; that is, $\mathbf{u} \ge \mathbf{0}$. Then $f(\mathbf{x}) \ge \theta(\mathbf{u}, \mathbf{v})$.

Proof

By the definition of θ, and since $\mathbf{x} \in X$, we have

$$\begin{aligned}\theta(\mathbf{u},\mathbf{v}) &= \inf\{f(\mathbf{y}) + \mathbf{u}^t\mathbf{g}(\mathbf{y}) + \mathbf{v}^t\mathbf{h}(\mathbf{y}) : \mathbf{y} \in X\} \\ &\le f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \le f(\mathbf{x})\end{aligned}$$

since $\mathbf{u} \ge \mathbf{0}$, $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$, and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. This completes the proof.

Corollary 1

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in X,\ \mathbf{g}(\mathbf{x}) \le \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0}\} \ge \sup\{\theta(\mathbf{u},\mathbf{v}) : \mathbf{u} \ge \mathbf{0}\}.$$

Corollary 2

If $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, where $\bar{\mathbf{u}} \ge \mathbf{0}$ and $\bar{\mathbf{x}} \in \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}) \le \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, then $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solve the primal and dual problems, respectively.

Corollary 3

If $\inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\} = -\infty$, then $\theta(\mathbf{u}, \mathbf{v}) = -\infty$ for each $\mathbf{u} \ge \mathbf{0}$.

Corollary 4

If $\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \ge \mathbf{0}\} = \infty$, then the primal problem has no feasible solution.

Duality Gap

From Corollary 1 to Theorem 6.2.1, the optimal objective value of the primal problem is greater than or equal to the optimal objective value of the dual problem. If strict inequality holds true, a *duality gap* is said to exist. Figure 6.3 illustrates the case of a duality gap for a problem having a single inequality constraint and no equality constraints. The perturbation function $v(y)$ for $y \in R$ is as shown in the figure. Note that by definition, this is the *greatest monotone nonincreasing function that envelopes G from below* (see Exercise 6.1). The optimal primal value is $v(0)$. The greatest intercept on the ordinate z-axis achieved by a hyperplane that supports G from below gives the optimal dual objective value as shown. In particular, observe that there does not exist a $\bar{u}$ such that $v(y) \ge v(0) - \bar{u}y$ for all $y \in R$, as we had in Figures 6.1 and 6.2. Exercise 6.2 asks the reader to construct G and v for the instance illustrated in Figure 4.13 that results in a situation similar to that of Figure 6.3.

6.2.2 Example

Consider the following problem:

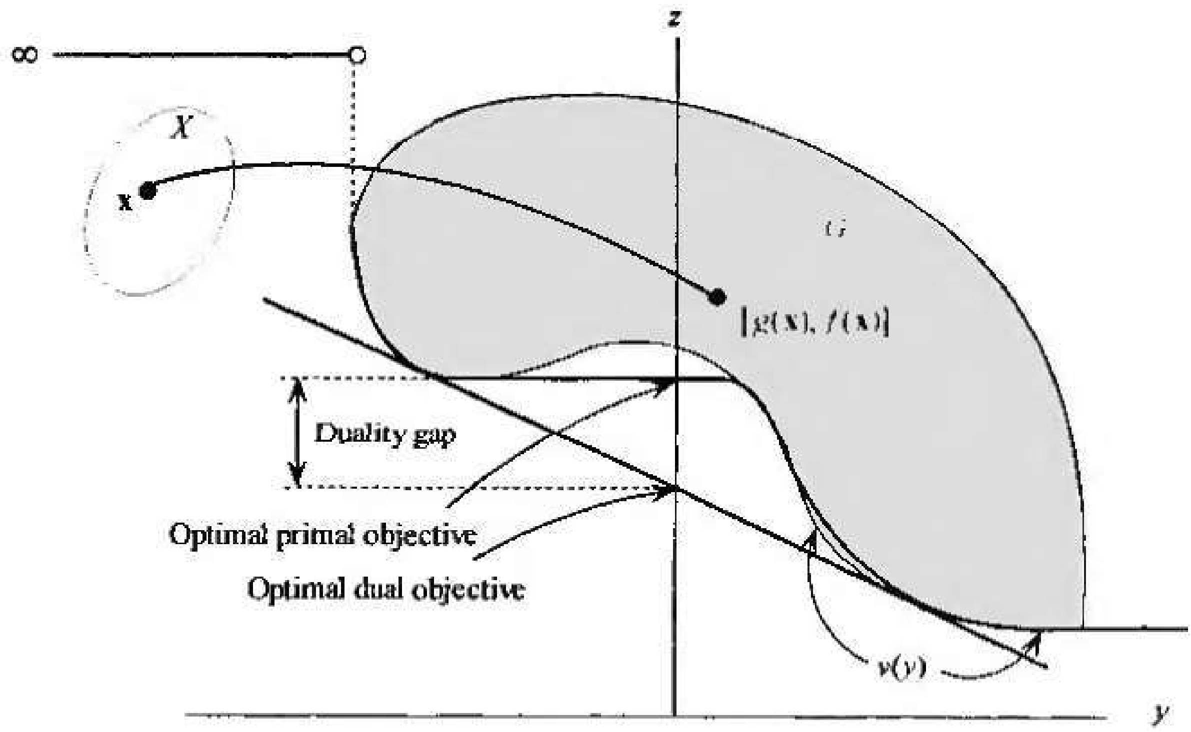


Figure 6.3 Duality gap.

$$\begin{aligned} \text{Minimize} \quad & f(x) = -2x_1 + x_2 \\ \text{subject to} \quad & h(x) = x_1 + x_2 - 3 = 0 \\ & (x_1, x_2) \in X, \end{aligned}$$

where $X = \{(0, 0), (0, 4), (4, 4), (4, 0), (1, 2), (2, 1)\}$.

It is easy to verify that (2, 1) is the optimal solution to the primal problem with objective value equal to –3. The dual objective function θ is given by

$$\theta(v) = \min\{(-2x_1 + x_2) + v(x_1 + x_2 - 3) : (x_1, x_2) \in X\}.$$

The reader may verify that the explicit expression for θ is given by

$$\theta(v) = \begin{cases} -4 + 5v & \text{for } v \le -1 \\ -8 + v & \text{for } -1 \le v \le 2 \\ -3v & \text{for } v \ge 2. \end{cases}$$

The dual function is shown in Figure 6.4, and the optimal solution is $\bar{v} = 2$ with objective value –6. Note that there exists a duality gap in this example.

In this case, the set G consists of a finite number of points, each corresponding to a point in X. This is shown in Figure 6.5. The supporting hyperplane, whose intercept on the vertical axis is maximal, is shown in the figure. Note that the intercept is equal to –6 and that the slope is equal to –2. Thus, the optimal dual solution is $\bar{v} = 2$, with objective value –6. Furthermore, note that the points in the set G on the vertical axis correspond to the primal feasible points and, hence, the minimal primal objective value is equal to –3.

Similar to the inequality constrained case, the perturbation function here is defined as $v(y) = \min\{f(\mathbf{x}) : h(\mathbf{x}) = y, \mathbf{x} \in X\}$. Because of the discrete nature of X, $h(\mathbf{x})$ can take on only a finite possible number of values. Hence,

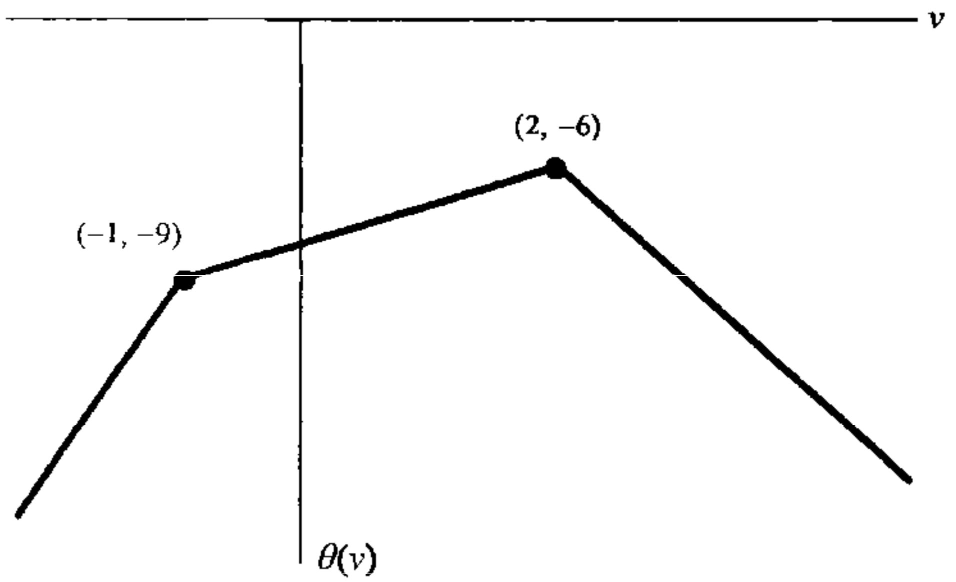


Figure 6.4 Dual function for Example 6.2.2.

noting G in Figure 6.5, we obtain $v(-3) = 0$, $v(0) = -3$, $v(1) = -8$, and $v(5) = -4$, with $v(y) = \infty$ for all $y \in R$ otherwise. Again, the optimal primal value is $v(0) = -3$, and there does not exist a $\bar{v}$ such that $v(y) \ge v(0) - \bar{v}y$. Hence, a duality gap exists.

Conditions that guarantee the absence of a duality gap are given in Theorem 6.2.4. Then Theorem 6.2.7 relates this to the perturbation function. First, however, the following lemma is needed.

6.2.3 Lemma

Let X be a nonempty convex set in R^n. Let α: $R^n \to R$ and g: $R^n \to R^m$ be convex, and let **h**: $R^n \to R^\ell$ be affine; that is, **h** is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{Ax} - \mathbf{b}$. If System 1 below has no solution **x**, then System 2 has a solution $(u_0, \mathbf{u}, \mathbf{v})$. The converse holds true if $u_0 > 0$.

$$\textit{System 1}: \quad \alpha(\mathbf{x}) < 0, \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \quad \mathbf{h}(\mathbf{x}) = \mathbf{0} \quad \text{for some } \mathbf{x} \in X$$

$$\textit{System 2}: \quad u_0\alpha(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \ge 0 \quad \text{for all } \mathbf{x} \in X$$
$$(u_0, \mathbf{u}) \ge \mathbf{0}, \quad (u_0, \mathbf{u}, \mathbf{v}) \ne \mathbf{0}.$$

Proof

Suppose that System 1 has no solution, and consider the following set:

$$\Lambda = \{(p, \mathbf{q}, \mathbf{r}) : p > \alpha(\mathbf{x}),\ \mathbf{q} \ge \mathbf{g}(\mathbf{x}),\ \mathbf{r} = \mathbf{h}(\mathbf{x}) \text{ for some } \mathbf{x} \in X\}.$$

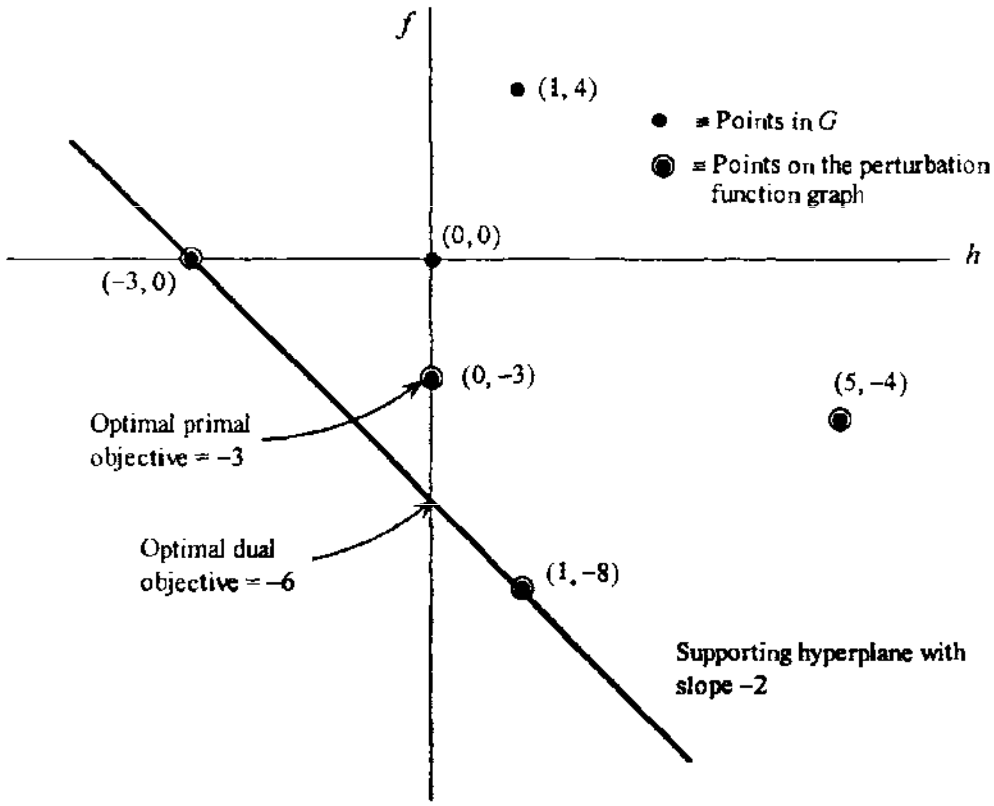


Figure 6.5 Geometric interpretation of Example 6.2.2.

Noting that X, α, and $\mathbf{g}$ are convex and that $\mathbf{h}$ is affine, it can easily be shown that Λ is convex. Since System 1 has no solution, $(0, \mathbf{0}, \mathbf{0}) \notin \Lambda$. By Corollary 1 to Theorem 2.4.7, there exists a nonzero $(u_0, \mathbf{u}, \mathbf{v})$ such that

$$u_0 p + \mathbf{u}^t \mathbf{q} + \mathbf{v}^t \mathbf{r} \geq 0 \qquad \text{for each } (p, \mathbf{q}, \mathbf{r}) \in \text{cl } \Lambda. \tag{6.1}$$

Now fix an $\mathbf{x} \in X$. Since p and $\mathbf{q}$ can be made arbitrarily large, (6.1) holds true only if $u_0 \geq 0$ and $\mathbf{u} \geq \mathbf{0}$. Furthermore, $(p, \mathbf{q}, \mathbf{r}) = [\alpha(\mathbf{x}), \mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x})]$ belongs to cl Λ. Therefore, from (6.1), we get

$$u_0\alpha(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \geq 0.$$

Since the above inequality is true for each $\mathbf{x} \in X$, System 2 has a solution.

To prove the converse, assume that System 2 has a solution $(u_0, \mathbf{u}, \mathbf{v})$ such that $u_0 > 0$ and $\mathbf{u} \geq \mathbf{0}$, satisfying

$$u_0\alpha(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \geq 0 \qquad \text{for each } \mathbf{x} \in X.$$

Now let $\mathbf{x} \in X$ be such that $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. From the above inequality, since $\mathbf{u} \geq \mathbf{0}$, we conclude that $u_0\alpha(\mathbf{x}) \geq 0$. Since $u_0 > 0$, $\alpha(\mathbf{x}) > 0$; and hence, System 1 has no solution. This completes the proof.

Theorem 6.2.4, referred to as the *strong duality theorem*, shows that under suitable convexity assumptions and under a constraint qualification, the optimal objective function values of the primal and dual problems are equal.

6.2.4 Theorem (Strong Duality Theorem)

Let X be a nonempty convex set in R^n, let f: $R^n \to R$ and $\mathbf{g}$: $R^n \to R^m$ be convex, and let $\mathbf{h}$: $R^n \to R^\ell$ be affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Suppose that the following constraint qualification holds true. There exists an $\hat{\mathbf{x}} \in X$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{0} \in \text{int } \mathbf{h}(X)$, where $\mathbf{h}(X) = \{\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}$. Then

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\} = \sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\}. \tag{6.2}$$

Furthermore, if the inf is finite, then $\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\}$ is achieved at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \geq \mathbf{0}$. If the inf is achieved at $\bar{\mathbf{x}}$, then $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$.

Proof

Let $\gamma = \inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$. By assumption, $\gamma < \infty$. If $\gamma = -\infty$, then by Corollary 3 to Theorem 6.2.1, $\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\} = -\infty$, and therefore (6.2) holds true. Hence, suppose that γ is finite, and consider the following system:

$$f(\mathbf{x})-\gamma<0, \quad \mathbf{g}(\mathbf{x})\leq\mathbf{0}, \quad \mathbf{h}(\mathbf{x})=\mathbf{0}, \quad \mathbf{x}\in X.$$

By the definition of γ, this system has no solution. Hence, from Lemma 6.2.3, there exists a nonzero vector $(u_0, \mathbf{u}, \mathbf{v})$ with $(u_0, \mathbf{u})\geq\mathbf{0}$ such that

$$u_0[f(\mathbf{x})-\gamma]+\mathbf{u}^t\mathbf{g}(\mathbf{x})+\mathbf{v}^t\mathbf{h}(\mathbf{x})\geq 0 \qquad \text{for all } \mathbf{x}\in X. \tag{6.3}$$

We first show that $u_0 > 0$. By contradiction, suppose that $u_0 = 0$. By assumption there exists an $\hat{\mathbf{x}}\in X$ such that $\mathbf{g}(\hat{\mathbf{x}})<\mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}})=\mathbf{0}$. Substituting in (6.3), it follows that $\mathbf{u}^t\mathbf{g}(\hat{\mathbf{x}})\geq 0$. Since $\mathbf{g}(\hat{\mathbf{x}})<\mathbf{0}$ and $\mathbf{u}\geq\mathbf{0}$, $\mathbf{u}^t\mathbf{g}(\hat{\mathbf{x}})\geq 0$ is possible only if $\mathbf{u} = \mathbf{0}$. But from (6.3), $u_0 = 0$ and $\mathbf{u} = \mathbf{0}$, which implies that $\mathbf{v}^t\mathbf{h}(\mathbf{x}) \geq 0$ for all $\mathbf{x}\in X$. But since $\mathbf{0}\in \operatorname{int} \mathbf{h}(X)$, we can pick an $\mathbf{x}\in X$ such that $\mathbf{h}(\mathbf{x}) = -\lambda\mathbf{v}$, where $\lambda > 0$. Therefore, $0\leq \mathbf{v}^t\mathbf{h}(\mathbf{x}) = -\lambda\|\mathbf{v}\|^2$, which implies that $\mathbf{v} = \mathbf{0}$. Thus, we have shown that $u_0 = 0$ implies that $(u_0, \mathbf{u}, \mathbf{v}) = \mathbf{0}$, which is impossible. Hence, $u_0 > 0$. Dividing (6.3) by u_0 and denoting $\mathbf{u}/u_0$ and $\mathbf{v}/u_0$ by $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$, respectively, we get

$$f(\mathbf{x})+\bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x})+\bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x})\geq\gamma \qquad \text{for all } \mathbf{x}\in X. \tag{6.4}$$

This shows that $\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})=\inf\{f(\mathbf{x})+\bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x})+\bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x}) : \mathbf{x}\in X\}\geq\gamma$. In view of Theorem 6.2.1, it is then clear that $\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})=\gamma$, and $(\bar{\mathbf{u}},\bar{\mathbf{v}})$ solves the dual problem.

To complete the proof, suppose that $\bar{\mathbf{x}}$ is an optimal solution to the primal problem; that is, $\bar{\mathbf{x}}\in X$, $\mathbf{g}(\bar{\mathbf{x}})\leq \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $f(\bar{\mathbf{x}})=\gamma$. From (6.4), letting $\mathbf{x}=\bar{\mathbf{x}}$, we get $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}})\geq 0$. Since $\bar{\mathbf{u}}\geq\mathbf{0}$ and $\mathbf{g}(\bar{\mathbf{x}})\leq\mathbf{0}$, we get $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}})=0$, and the proof is complete.

In Theorem 6.2.4, the assumption $\mathbf{0}\in \operatorname{int}\mathbf{h}(X)$ and that there exists an $\hat{\mathbf{x}}\in X$ such that $\mathbf{g}(\hat{\mathbf{x}})<\mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}})=\mathbf{0}$ can be viewed as a generalization of Slater's constraint qualification of Chapter 5. In particular, if $X = R^n$, then $\mathbf{0}\in \operatorname{int} \mathbf{h}(X)$ holds true automatically (if redundant equations are deleted), so that the constraint qualification asserts the existence of a point $\hat{\mathbf{x}}$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$. To see this, suppose that $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Without loss of generality, assume that rank $(\mathbf{A}) = m$, because otherwise, any redundant constraints could be deleted. Now, any $\mathbf{y}\in R^m$ could be represented as $\mathbf{y} = \mathbf{A}\mathbf{x} - \mathbf{b}$, where $\mathbf{x} = \mathbf{A}^t(\mathbf{A}\mathbf{A}^t)^{-1}(\mathbf{y}+\mathbf{b})$. Thus, $\mathbf{h}(X) = R^m$ and, in particular, $\mathbf{0}\in \operatorname{int}\mathbf{h}(X)$.

Saddle Point Criteria

The foregoing theorem shows that under convexity assumptions and under a suitable constraint qualification, the primal and dual objective function values match at optimality. Actually, a necessary and sufficient condition for the latter property to hold true is the existence of a saddle point, as we learn next. Given the primal Problem P, define the *Lagrangian function*

$$\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}).$$

A solution $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is called a *saddle point* of the Lagrangian function if $\bar{\mathbf{x}} \in X$, $\bar{\mathbf{u}} \geq \mathbf{0}$, and

$$\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \leq \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$$
$$\text{for all } \mathbf{x} \in X, \text{ and all } (\mathbf{u}, \mathbf{v}) \text{ with } \mathbf{u} \geq \mathbf{0}. \tag{6.5}$$

Hence, we have that $\bar{\mathbf{x}}$ minimizes ϕ over X when $(\mathbf{u}, \mathbf{v})$ is fixed at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, and that $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ maximizes ϕ over all $(\mathbf{u}, \mathbf{v})$ with $\bar{\mathbf{u}} \geq \mathbf{0}$ when $\mathbf{x}$ is fixed at $\bar{\mathbf{x}}$. Relating this to Figure 4.2, we see why $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is called a *saddle point* for the Lagrangian function ϕ.

The following result characterizes a saddle point solution and shows that its existence is a necessary and sufficient condition for the absence of a duality gap.

6.2.5 Theorem (Saddle Point Optimality and Absence of a Duality Gap)

A solution $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in X$ and $\bar{\mathbf{u}} \geq \mathbf{0}$ is a saddle point for the Lagrangian function $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$ if and only if

a. $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = \min\{\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) : \mathbf{x} \in X\}$,

b. $\mathbf{g}(\bar{\mathbf{x}}) \leq \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and

c. $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$.

Moreover, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point if and only if $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ are, respectively, optimal solutions to the primal and dual problems P and D with no duality gap, that is, with $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$.

Proof

Suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point for the Lagrangian function ϕ. By definition, Condition (a) must be true. Furthermore, from (6.5), we have

$$f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) \geq f(\bar{\mathbf{x}}) + \mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{v}^t\mathbf{h}(\bar{\mathbf{x}}) \quad \text{for all } (\mathbf{u}, \mathbf{v}) \text{ with } \mathbf{u} \geq \mathbf{0}. \tag{6.6}$$

Clearly, this implies that we must have $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$ and $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, or else (6.6) can be violated by appropriately making a component of $\mathbf{u}$ or $\mathbf{v}$ sufficiently large in magnitude. Now, taking $\mathbf{u} = \mathbf{0}$ in (6.6), we obtain that $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) \ge \mathbf{0}$. Noting that $\bar{\mathbf{u}} \ge \mathbf{0}$ and $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$ imply that $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) \le 0$, we must have $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$. Hence, conditions (a), (b), and (c) hold true.

Conversely, suppose that we are given $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in X$ and $\bar{\mathbf{u}} \ge \mathbf{0}$ such that conditions (a), (b), and (c) hold true. Then $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \le \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{x} \in X$ by Property (a). Furthermore, $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) \ge f(\bar{\mathbf{x}}) + \mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{v}^t\mathbf{h}(\bar{\mathbf{x}}) = \phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v})$ for all $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \ge \mathbf{0}$, since $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$ and $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$. Hence, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point. This proves the first part of the theorem.

Next, suppose again that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point. By Property (b), $\bar{\mathbf{x}}$ is feasible to Problem P. Since $\bar{\mathbf{u}} \ge \mathbf{0}$, we also have that $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ is feasible to D. Moreover, by properties (a), (b), and (c), $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}})$. By Corollary 2 to Theorem 6.2.1, $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solve P and D, respectively, with no duality gap.

Finally, suppose that $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ are optimal solutions to problems P and D, respectively, with $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Hence, we have $\bar{\mathbf{x}} \in X$, $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $\bar{\mathbf{u}} \ge \mathbf{0}$. Moreover, we have by primal–dual feasibility that

$$\begin{aligned}\theta(\bar{\mathbf{u}},\bar{\mathbf{v}}) &= \min\{f(\mathbf{x}) + \bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\} \\ &\le f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) \le f(\bar{\mathbf{x}}).\end{aligned}$$

But $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}})$, by hypothesis. Hence, equality holds true throughout the discussion above. In particular, $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$, so $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \min\{\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) : \mathbf{x} \in X\}$. Hence, properties (a), (b), and (c) hold true in addition to $\bar{\mathbf{x}} \in X$ and $\bar{\mathbf{u}} \ge \mathbf{0}$; so $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point. This completes the proof.

Corollary

Suppose that X, f, and $\mathbf{g}$ are convex and that $\mathbf{h}$ is affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Further, suppose that $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$ and that there exists an $\hat{\mathbf{x}} \in X$ with $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$. If $\bar{\mathbf{x}}$ is an optimal solution to the primal Problem P, there exists a vector $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \ge \mathbf{0}$ such that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point.

Proof

By Theorem 6.2.4 there exists an optimal solution $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, $\bar{\mathbf{u}} \geq \mathbf{0}$ to Problem D such that $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Hence, by Theorem 6.2.5, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point solution. This completes the proof.

There is an additional insight that can be derived in regard to the duality gap between the primal and dual problems. Note that the dual problem's optimal value is given by

$$\theta^* = \sup_{(\mathbf{u},\mathbf{v}):\mathbf{u}\geq \mathbf{0}} \inf_{\mathbf{x}\in X} [\phi(\mathbf{x},\mathbf{u},\mathbf{v})].$$

If we interchange the order of optimization (see Exercise 6.3), we get

$$\theta^* \leq \inf_{\mathbf{x}\in X} \sup_{(\mathbf{u},\mathbf{v}):\mathbf{u}\geq \mathbf{0}} [\phi(\mathbf{x},\mathbf{u},\mathbf{v})].$$

But the supremum of $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$ over $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \geq \mathbf{0}$ is infinity, unless $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, whence it is $f(\mathbf{x})$. Hence,

$$\begin{aligned}\theta^* &\leq \inf_{\mathbf{x}\in X} \sup_{(\mathbf{u},\mathbf{v}):\mathbf{u}\geq \mathbf{0}} [\phi(\mathbf{x},\mathbf{u},\mathbf{v})] \\ &= \inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{x} \in X\},\end{aligned}$$

which is the primal optimal value. Hence, we see that the primal and dual objective values match at optimality if and only if the interchange of the foregoing infimum and supremum operations leaves the optimal values unchanged. By Theorem 6.2.5, assuming that an optimum exists, this occurs if and only if there exists a saddle point $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for the Lagrangian function ϕ.

Relationship Between the Saddle Point Criteria and the Karush–Kuhn–Tucker Conditions

In Chapters 4 and 5, we discussed the KKT optimality conditions for Problem P:

$$\begin{aligned}\text{Minimize } & f(\mathbf{x}) \\ \text{subject to } & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in X.\end{aligned}$$

Furthermore, in Theorem 6.2.5 we developed the saddle point optimality conditions for the same problem. Theorem 6.2.6 gives the relationship between these two types of optimality conditions.

6.2.6 Theorem

Let $S = \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}) \le \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, and consider Problem P to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Suppose that $\bar{\mathbf{x}} \in S$ satisfies the KKT conditions; that is, there exist $\bar{\mathbf{u}} \ge \mathbf{0}$ and $\bar{\mathbf{v}}$ such that

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) + \nabla \mathbf{g}(\bar{\mathbf{x}})^t \bar{\mathbf{u}} + \nabla \mathbf{h}(\bar{\mathbf{x}})^t \bar{\mathbf{v}} &= \mathbf{0} \\ \bar{\mathbf{u}}^t \mathbf{g}(\bar{\mathbf{x}}) &= 0. \end{aligned} \tag{6.7}$$

Suppose that f and g_i for $i \in I$ are convex at $\bar{\mathbf{x}}$, where $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Further, suppose that if $\bar{v}_i \ne 0$, then h_i is affine. Then $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point for the Lagrangian function $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t \mathbf{g}(\mathbf{x}) + \mathbf{v}^t \mathbf{h}(\mathbf{x})$.

Conversely, suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in \operatorname{int} X$ and $\bar{\mathbf{u}} \ge 0$ is a saddle point solution. Then $\bar{\mathbf{x}}$ is feasible to Problem P, and furthermore, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ satisfies the KKT conditions specified by (6.7).

Proof

Suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$, with $\bar{\mathbf{x}} \in S$ and $\bar{\mathbf{u}} \ge \mathbf{0}$, satisfies the KKT conditions specified by (6.7). By convexity at $\bar{\mathbf{x}}$ of f and g_i for $i \in I$, and since h_i is affine for $\bar{v}_i \ne 0$, we get

$$f(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}) \tag{6.8a}$$

$$g_i(\mathbf{x}) \ge g_i(\bar{\mathbf{x}}) + \nabla g_i(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for } i \in I \tag{6.8b}$$

$$h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}}) + \nabla h_i(\bar{\mathbf{x}})^t (\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for } i = 1,\ldots,\ell,\ \bar{v}_i \ne 0 \tag{6.8c}$$

for all $\mathbf{x} \in X$. Multiplying (6.8b) by $\bar{u}_i \ge 0$, (6.8c) by $\bar{v}_i$, and adding these to (6.8a) and noting (6.7), it follows from the definition of ϕ that $\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \ge \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{x} \in X$. Also, since $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $\bar{\mathbf{u}}^t \mathbf{g}(\bar{\mathbf{x}}) = 0$, it follows that $\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \le \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \ge \mathbf{0}$. Hence, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ satisfies the saddle point conditions given by (6.5).

To prove the converse, suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in \operatorname{int} X$ and $\bar{\mathbf{u}} \ge \mathbf{0}$ is a saddle point solution. Since $\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \le \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{u} \ge \mathbf{0}$ and all $\mathbf{v}$, we have, using (6.6) as in Theorem 6.2.5, that $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $\bar{\mathbf{u}}^t \mathbf{g}(\bar{\mathbf{x}}) = 0$. This shows that $\bar{\mathbf{x}}$ is feasible to Problem P. Since $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \le \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{x} \in X$, then $\bar{\mathbf{x}}$ solves the problem to minimize $\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ subject to $\mathbf{x} \in X$. Since $\bar{\mathbf{x}} \in \operatorname{int} X$, then $\nabla_{\mathbf{x}} \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = \mathbf{0}$, that is, $\nabla f(\bar{\mathbf{x}}) + \nabla \mathbf{g}(\bar{\mathbf{x}})^t \bar{\mathbf{u}} + \nabla \mathbf{h}(\bar{\mathbf{x}})^t \bar{\mathbf{v}} = \mathbf{0}$; and hence, (6.7) holds true. This completes the proof.

Theorem 6.2.6 shows that if $\bar{x}$ is a KKT point, then under certain convexity assumptions, the Lagrangian multipliers in the KKT conditions also serve as the multipliers in the saddle point criteria. Conversely, the multipliers in the saddle point conditions are the Lagrangian multipliers of the KKT conditions. Moreover, in view of Theorems 6.2.4, 6.2.5, and 6.2.6, the optimal dual variables for the Lagrangian dual problem are precisely the Lagrangian multipliers for the KKT conditions and also the multipliers for the saddle point conditions in this case.

Saddle Point Optimality Interpretation Using a Perturbation Function

While discussing the geometric interpretation of the dual problem and the associated duality gap, we introduced the concept of a perturbation function v and illustrated this in Examples 6.1.1 and 6.1.2 (see Figures 6.1 through 6.5). As alluded to previously, the existence of a supporting hyperplane to the epigraph of this function at the point (0, $v(0)$) related to the absence of a duality gap in these examples. This is formalized in the discussion that follows.

Consider the primal Problem P, and define the *perturbation function* v: $R^{m+\ell} \to R$ as the optimal value function of the following problem, where $\mathbf{y} = (y_1,\ldots,y_m,y_{m+1},\ldots,y_{m+\ell})$:

$$v(\mathbf{y}) = \min\{f(\mathbf{x}) : g_i(\mathbf{x}) \le y_i \quad \text{for } i = 1,\ldots, m,$$

$$h_i(\mathbf{x}) = y_{m+i} \quad \text{for } i = 1,\ldots,\ell,\ \mathbf{x} \in X\}. \qquad (6.9)$$

Theorem 6.2.7 asserts that if Problem P has an optimum, the existence of a saddle point solution, that is, the absence of a duality gap, is equivalent to the existence of a supporting hyperplane for the epigraph of v at the point $(0, v(\mathbf{0}))$.

6.2.7 Theorem

Consider the primal Problem P, and assume that an optimal solution $\bar{\mathbf{x}}$ to this problem exists. Then $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point for the Lagrangian function $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$ if and only if

$$v(\mathbf{y}) \ge v(0) - (\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)\mathbf{y} \qquad \text{for all } \mathbf{y} \in R^{m+\ell}, \qquad (6.10)$$

that is, if and only if the hyperplane $z = v(\mathbf{0}) - (\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)\mathbf{y}$ supports the epigraph $\{(\mathbf{y}, z) : z \ge v(\mathbf{y}),\ \mathbf{y} \in R^{m+\ell}\}$ of v at the point $(\mathbf{y}, z) = (0, v(\mathbf{0}))$.

Proof

Suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point solution. Then, by Theorem 6.2.5, the absence of a duality gap asserts that

$$v(\mathbf{0}) = \theta(\bar{\mathbf{u}},\bar{\mathbf{v}}) = \min\{f(\mathbf{x})+\bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x})+\bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x}) : \mathbf{x}\in X\}$$

$$= (\bar{\mathbf{u}}^t,\bar{\mathbf{v}}^t)\mathbf{y} + \min\left\{ f(\mathbf{x}) + \sum_{i=1}^{m}\bar{u}_i[g_i(\mathbf{x})-y_i] + \sum_{i=1}^{\ell}\bar{v}_i[h_i(\mathbf{x})-y_{m+i}] : \mathbf{x}\in X\right\} \qquad \text{for any } \mathbf{y}\in R^{m+\ell}.$$

Applying the weak duality Theorem 6.2.1 to the perturbed problem (6.9), we obtain from the foregoing identity that $v(0) \le (\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)\mathbf{y} + v(\mathbf{y})$ for any $\mathbf{y} \in R^{m+\ell}$, so (6.10) holds true.

Conversely, suppose that (6.10) holds true for some $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, and let $\bar{\mathbf{x}}$ solve Problem P. We must show that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point solution. First, note that $\bar{\mathbf{x}} \in X$, $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, and $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$. Furthermore, $\bar{\mathbf{u}} \ge \mathbf{0}$ must hold true, because if $\bar{u}_p < 0$, say, then by selecting y such that $y_i = 0$ for $i \ne p$, and $y_p > 0$, we obtain $v(0) \ge v(\mathbf{y}) \ge v(0) - \bar{u}_p y_p$, which implies that $\bar{u}_p y_p \ge 0$, a contradiction.

Second, observe that by fixing $\mathbf{y} = \bar{\mathbf{y}} \equiv [\mathbf{g}(\bar{\mathbf{x}})^t,\ \mathbf{h}(\bar{\mathbf{x}})^t]$ in (6.9), we obtain a restriction of Problem P, since $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$ and $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$. But for the same reason, since $\bar{\mathbf{x}}$ is feasible to (6.9) with y fixed as such, and since $\bar{\mathbf{x}}$ solves Problem P, we obtain $v(\bar{\mathbf{y}}) = v(\mathbf{0})$. By (6.10), this in turn means that $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) \ge 0$. Since $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$ and $\bar{\mathbf{u}} \ge 0$, we therefore must have $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$.

Finally, we have

$$\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) = v(\mathbf{0}) \le v(\mathbf{y}) + (\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)\mathbf{y} \qquad (6.11)$$

for all $\mathbf{y} \in R^{m+\ell}$. Now, for any $\hat{\mathbf{x}} \in X$, denoting $\hat{\mathbf{y}} = [\mathbf{g}(\hat{\mathbf{x}})^t, \mathbf{h}(\hat{\mathbf{x}})^t]$, we obtain from (6.9) that $v(\hat{\mathbf{y}}) \le f(\hat{\mathbf{x}})$, since $\hat{\mathbf{x}}$ is feasible to (6.9) with $\mathbf{y} = \hat{\mathbf{y}}$. Hence, using this in (6.11), we obtain $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\hat{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\hat{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\hat{\mathbf{x}})$ all $\hat{\mathbf{x}} \in X$; so $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = \min\{\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) : \mathbf{x} \in X\}$.

We have therefore shown that $\bar{\mathbf{x}} \in X$, $\bar{\mathbf{u}} \ge \mathbf{0}$, and that conditions (a), (b), and (c) of Theorem 6.2.5 hold true. Consequently, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point for ϕ, and this completes the proof.

To illustrate, observe in Figures 6.1 and 6.2 that there does exist a supporting hyperplane for the epigraph of v at $(0, v(0))$. Hence, both the primal and dual problems have optimal solutions having the same optimal objective values for these cases. However, for the situations illustrated in Figures 6.3 and 6.5, no such supporting hyperplane exists. Hence, these instances possess a positive duality gap.

In conclusion, there are two noteworthy points regarding the perturbation function v. First, if f and $\mathbf{g}$ are convex, $\mathbf{h}$ is affine, and X is a convex set, it can easily be shown that v is a convex function (see Exercise 6.4). Hence, in this case, the condition (6.10) reduces to the statement that $-(\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)$ is a subgradient of v at $\mathbf{y} = \mathbf{0}$.

Second, suppose that corresponding to the primal and dual problems P and D there exists a saddle point solution $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$, and assume that v is continuously differentiable at $\mathbf{y} = \mathbf{0}$. Then we have $v(\mathbf{y}) = v(\mathbf{0}) + \nabla v(\mathbf{0})^t \mathbf{y} + \|\mathbf{y}\|\alpha(\mathbf{0};\mathbf{y})$, where $\alpha(\mathbf{0};\mathbf{y}) \to 0$ as $\mathbf{y} \to \mathbf{0}$. Using (6.10) of Theorem 6.2.7, this means that $-[\nabla v(\mathbf{0})^t + (\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)]\mathbf{y} \le \|\mathbf{y}\|\alpha(\mathbf{0};\mathbf{y})$ for all $\mathbf{y} \in R^{m+\ell}$. Letting $\mathbf{y} = -\lambda[\nabla v(\mathbf{0})^t + (\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)]$ for $\lambda \ge 0$, and letting $\lambda \to 0^+$, we readily conclude that $\nabla v(\mathbf{0})^t = -(\bar{\mathbf{u}}^t, \bar{\mathbf{v}}^t)$. Hence, the negative of the optimal Lagrange multiplier values give the *marginal rates of change* in the optimal objective value of Problem P with respect to perturbations in the right-hand sides. Assuming that the problem represents one of minimizing cost subject to various material, labor, budgetary resource limitations, and demand constraints, this yields useful *economic interpretations* in terms of the marginal change in the optimal cost with respect to perturbations in such resource or demand entities.

6.2.8 Example

Consider the following (primal) problem:

$$\text{P:} \quad \text{Minimize}\{x_2 : x_1 \ge 1,\ x_1^2 + x_2^2 \le 1, (x_1, x_2) \in R^2\}.$$

As illustrated in Figure 6.6, the unique optimal solution to this problem is $(\bar{x}_1, \bar{x}_2) = (1, 0)$, with optimal objective function value equal to 0. However, although this is a convex programming problem, the optimum is not a KKT point, since $F_0 \cap G' \ne \varnothing$, and a saddle point solution does not exist (refer to Theorems 4.2.15 and 6.2.6).

Now let us formulate the Lagrangian dual Problem D by treating $1 - x_1 \le 0$ as $g(\mathbf{x}) \le 0$ and letting X denote the set $\{(x_1, x_2) : x_1^2 + x_2^2 \le 1\}$. Hence, Problem D requires us to find $\sup\{\theta(u) : u \ge 0\}$, where $\theta(u) = \inf\{x_2 + u(1 - x_1) : x_1^2 + x_2^2 \le 1\}$. For any $u \ge 0$, it is readily verified that the optimum is attained at $x_1 = u/\sqrt{1+u^2}$ and $x_2 = -1/\sqrt{1+u^2}$. Hence, $\theta(u) = u - \sqrt{1+u^2}$. We see that as $u \to \infty$, $\theta(u) \to 0$, the optimal primal objective value. Hence, $\sup\{\theta(u) : u \ge 0\} = 0$, but this is not attained for any $\bar{u} \ge 0$; that is, a maximizing solution $\bar{u}$ does not exist.

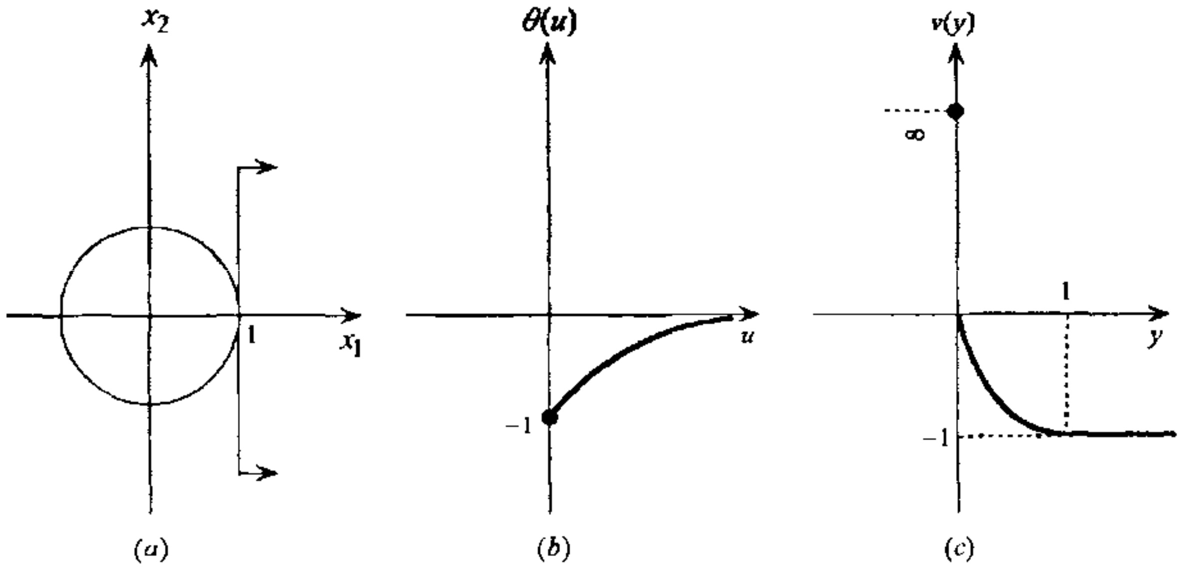


Figure 6.6 Solution to Example 6.2.8.

Next, let us determine the perturbation function $v(y)$ for $y \in R$. Note that $v(y) = \min\{x_2 : 1 - x_1 \le y,\ x_1^2 + x_2^2 \le 1\}$. Hence, we obtain $v(y) = \infty$ for $y < 0$, $v(y) = -\sqrt{y(2-y)}$ for $0 \le y \le 1$, and $v(y) = -1$ for $y \ge 1$. This is illustrated in Figure 6.6*c*. Observe that there does not exist any supporting hyperplane at (0, 0) for the epigraph of $v(y)$, $y \in R$, since the right-hand derivative of v with respect to y at $y = 0$ is $-\infty$.

6.3 Properties of the Dual Function

In Section 6.2 we studied the relationships between the primal and dual problems. Under certain conditions, Theorems 6.2.4 and 6.2.5 showed that the optimal objective values of the primal and dual problems are equal, and hence it would be possible to solve the primal problem indirectly by solving the dual problem. To facilitate solution of the dual problem, we need to examine the properties of the dual function. In particular, we show that θ is concave, discuss its differentiability and subdifferentiability properties, and characterize its ascent and steepest ascent directions.

Throughout the rest of this chapter, we assume that the set X is compact. This will simplify the proofs of several of the theorems. Note that this assumption is not unduly restrictive, since if X were not bounded, we could add suitable lower and upper bounds on the variables such that the feasible region is not affected in the relative vicinity of an optimum. For convenience, we also combine the vectors $\mathbf{u}$ and $\mathbf{v}$ as $\mathbf{w}$ and the functions $\mathbf{g}$ and $\mathbf{h}$ as $\boldsymbol{\beta}$. Theorem 6.3.1 shows that θ is concave.

6.3.1 Theorem

Let X be a nonempty compact set in R^n, and let f: $R^n \to R$ and $\boldsymbol{\beta}$: $R^n \to R^{m+\ell}$ be continuous. Then θ, defined by

$$\theta(\mathbf{w}) = \inf\{f(\mathbf{x}) + \mathbf{w}^t \boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\},$$

is concave over $R^{m+\ell}$.

Proof

Since f and β are continuous and X is compact, θ is finite everywhere on $R^{m+\ell}$. Let $\mathbf{w}_1$, $\mathbf{w}_2 \in R^{m+\ell}$, and let $\lambda \in (0,1)$. We then have

$$\begin{aligned}\theta[\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2] &= \inf\{f(\mathbf{x}) + [\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2]^t \boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\ &= \inf\{\lambda[f(\mathbf{x}) + \mathbf{w}_1^t \boldsymbol{\beta}(\mathbf{x})] + (1-\lambda)[f(\mathbf{x}) + \mathbf{w}_2^t \boldsymbol{\beta}(\mathbf{x})] : \mathbf{x} \in X\} \\ &\geq \lambda \inf\{f(\mathbf{x}) + \mathbf{w}_1^t \boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\ &\quad + (1-\lambda) \inf\{f(\mathbf{x}) + \mathbf{w}_2^t \boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\ &= \lambda\theta(\mathbf{w}_1) + (1-\lambda)\theta(\mathbf{w}_2).\end{aligned}$$

Thus, θ is concave, and the proof is complete.

Since θ is concave, by Theorem 3.4.2, a local optimum of θ is also a global optimum. This makes the maximization of θ an attractive proposition. However, the main difficulty in solving the dual problem is that the dual function is not explicitly available, since θ can be evaluated at a point only after a minimization subproblem is solved. In the remainder of this section we study differentiability and subdifferentiability properties of the dual function. These properties will aid us in maximizing the dual function.

Differentiability of θ

We now address the question of differentiability of θ, defined by $\theta(\mathbf{w}) = \inf\{f(\mathbf{x}) + \mathbf{w}^t \boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\}$. It will be convenient to introduce the following set of optimal solutions to the Lagrangian dual subproblem:

$$X(\mathbf{w}) = \{\mathbf{y} : \mathbf{y} \text{ minimizes } f(\mathbf{x}) + \mathbf{w}^t \boldsymbol{\beta}(\mathbf{x}) \text{ over } \mathbf{x} \in X\}.$$

The differentiability of θ at any given point $\bar{\mathbf{w}}$ depends on the elements of $X(\bar{\mathbf{w}})$. In particular, if the set $X(\bar{\mathbf{w}})$ is a singleton, Theorem 6.3.3 shows that θ is differentiable at $\bar{\mathbf{w}}$. First, however, the following lemma is needed.

6.3.2 Lemma

Let X be a nonempty compact set in R^n, and let f: $R^n \to R$ and β: $R^n \to R^{m+\ell}$ be continuous. Let $\bar{\mathbf{w}} \in R^{m+\ell}$, and suppose that $X(\bar{\mathbf{w}})$ is the singleton $\{\bar{\mathbf{x}}\}$. Suppose that $\mathbf{w}_k \to \bar{\mathbf{w}}$, and let $\mathbf{x}_k \in X(\mathbf{w}_k)$ for each k. Then $\mathbf{x}_k \to \bar{\mathbf{x}}$.

Proof

By contradiction, suppose that $w_k \to \bar{w}$, $x_k \in X(w_k)$, and $\|x_k - \bar{x}\| > \varepsilon > 0$ for all $k \in \mathscr{K}$ where $\mathscr{K}$ is some index set. Since X is compact, the sequence $\{x_k\}_{\mathscr{K}}$ has a convergent subsequence $\{x_k\}_{\mathscr{K}'}$, with limit y in X. Note that $\|y - \bar{x}\| \geq \varepsilon > 0$, and hence, y and $\bar{x}$ are distinct. Furthermore, for each w_k with $k \in \mathscr{K}'$, we have

$$f(x_k) + w_k^t \beta(x_k) \leq f(\bar{x}) + w_k^t \beta(\bar{x}).$$

Taking the limit as k in $\mathscr{K}'$ approaches ∞, and noting that $x_k \to y$, $w_k \to \bar{w}$, and that f and β are continuous, it follows that

$$f(y) + \bar{w}^t \beta(y) \leq f(\bar{x}) + \bar{w}^t \beta(\bar{x}).$$

Therefore, $y \in X(\bar{w})$, contradicting the assumption that $X(\bar{w})$ is a singleton. This completes the proof.

6.3.3 Theorem

Let X be a nonempty compact set in R^n, and let f: $R^n \to R$, and β: $R^n \to R^{m+\ell}$ be continuous. Let $\bar{w} \in R^{m+\ell}$ and suppose that $X(\bar{w})$ is the singleton $\{\bar{x}\}$. Then θ is differentiable at $\bar{w}$ with gradient $\nabla\theta(\bar{w}) = \beta(\bar{x})$.

Proof

Since f and β are continuous and X is compact, for any given w there exists an $x_w \in X(w)$. From the definition of θ, the following two inequalities hold true:

$$\theta(w) - \theta(\bar{w}) \leq f(\bar{x}) + w^t \beta(\bar{x}) - f(\bar{x}) - \bar{w}^t \beta(\bar{x}) = (w - \bar{w})^t \beta(\bar{x}). \tag{6.12}$$

$$\theta(\bar{w}) - \theta(w) \leq f(x_w) + \bar{w}^t \beta(x_w) - f(x_w) - w^t \beta(x_w) = (\bar{w} - w)^t \beta(x_w). \tag{6.13}$$

From (6.12) and (6.13) and the Schwartz inequality, it follows that

$$\begin{aligned} 0 \geq \theta(w) - \theta(\bar{w}) - (w - \bar{w})^t \beta(\bar{x}) &\geq (w - \bar{w})^t [\beta(x_w) - \beta(\bar{x})] \\ &\geq -\|w - \bar{w}\| \|\beta(x_w) - \beta(\bar{x})\|. \end{aligned}$$

This further implies that

$$0 \geq \frac{\theta(w) - \theta(\bar{w}) - (w - \bar{w})^t \beta(\bar{x})}{\|w - \bar{w}\|} \geq -\|\beta(x_w) - \beta(\bar{x})\|. \tag{6.14}$$

As $\mathbf{w} \to \overline{\mathbf{w}}$, then, by Lemma 6.3.2, $\mathbf{x}_w \to \overline{\mathbf{x}}$ and by the continuity of $\boldsymbol{\beta}$, $\boldsymbol{\beta}(\mathbf{x}_w) \to \boldsymbol{\beta}(\overline{\mathbf{x}})$. Therefore, from (6.14) we get

$$\lim_{\mathbf{w}\to\overline{\mathbf{w}}} \frac{\theta(\mathbf{w}) - \theta(\overline{\mathbf{w}}) - (\mathbf{w} - \overline{\mathbf{w}})^t \boldsymbol{\beta}(\overline{\mathbf{x}})}{\|\mathbf{w} - \overline{\mathbf{w}}\|} = 0.$$

Hence, θ is differentiable at $\overline{\mathbf{w}}$ with gradient $\boldsymbol{\beta}(\overline{\mathbf{x}})$. This completes the proof.

Subgradients of θ

We have shown in Theorem 6.3.1 that θ is concave, and hence, by Theorem 3.2.5, θ is subdifferentiable; that is, it has subgradients. As will be shown later, subgradients play an important role in the maximization of the dual function, since they lead naturally to the characterization of the directions of ascent. Theorem 6.3.4 shows that each $\overline{\mathbf{x}} \in X(\overline{\mathbf{w}})$ yields a subgradient of θ at $\overline{\mathbf{w}}$.

6.3.4 Theorem

Let X be a nonempty compact set in R^n, and let f: $R^n \to R$ and $\boldsymbol{\beta}$: $R^n \to R^{m+\ell}$ be continuous so that for any $\overline{\mathbf{w}} \in R^{m+\ell}$, $X(\overline{\mathbf{w}})$ is not empty. If $\overline{\mathbf{x}} \in X(\overline{\mathbf{w}})$, then $\boldsymbol{\beta}(\overline{\mathbf{x}})$ is a subgradient of θ at $\overline{\mathbf{w}}$.

Proof

Since f and $\boldsymbol{\beta}$ are continuous and X is compact, $X(\overline{\mathbf{w}}) \neq \varnothing$ for any $\overline{\mathbf{w}} \in R^{m+\ell}$. Now, let $\overline{\mathbf{w}} \in R^{m+\ell}$, and let $\overline{\mathbf{x}} \in X(\overline{\mathbf{w}})$. Then

$$\begin{aligned}\theta(\mathbf{w}) &= \inf\{f(\mathbf{x}) + \mathbf{w}^t \boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\ &\le f(\overline{\mathbf{x}}) + \mathbf{w}^t \boldsymbol{\beta}(\overline{\mathbf{x}}) \\ &= f(\overline{\mathbf{x}}) + (\mathbf{w} - \overline{\mathbf{w}})^t \boldsymbol{\beta}(\overline{\mathbf{x}}) + \overline{\mathbf{w}}^t \boldsymbol{\beta}(\overline{\mathbf{x}}) \\ &= \theta(\overline{\mathbf{w}}) + (\mathbf{w} - \overline{\mathbf{w}})^t \boldsymbol{\beta}(\overline{\mathbf{x}}).\end{aligned}$$

Therefore, $\boldsymbol{\beta}(\overline{\mathbf{x}})$ is a subgradient of θ at $\overline{\mathbf{w}}$, and the proof is complete.

6.3.5 Example

Consider the following primal problem:

$$\begin{aligned}\text{Minimize} \quad & -x_1 - x_2 \\ \text{subject to} \quad & x_1 + 2x_2 - 3 \le 0 \\ & x_1, x_2 = 0, 1, 2, \text{ or } 3.\end{aligned}$$

Letting $g(x_1, x_2) = x_1 + 2x_2 - 3$ and $X = \{(x_1, x_2) : x_1, x_2 = 0, 1, 2, \text{ or } 3\}$, the dual function is given by

$$\theta(u) = \inf\,\{-x_1 - x_2 + u(x_1 + 2x_2 - 3) : x_1, x_2 = 0, 1, 2, \text{ or } 3\}$$

$$= \begin{cases} -6+6u & \text{if } 0 \le u \le 1/2 \\ -3 & \text{if } 1/2 \le u \le 1 \\ -3u & \text{if } u \ge 1. \end{cases}$$

We ask the reader to plot the perturbation function for this example in Exercise 6.5 and to investigate the saddle point optimality conditions. Now let $\bar{u}$ = 1/2. To find a subgradient of θ at $\bar{u}$, consider the following subproblem:

$$\begin{aligned} &\text{Minimize} \quad -x_1 - x_2 + (1/2)(x_1 + 2x_2 - 3) \\ &\text{subject to} \quad x_1, x_2 = 0,\ 1,\ 2,\ \text{or}\ 3. \end{aligned}$$

Note that the set $X(\bar{u})$ of optimal solutions to the above problem is $\{(3, 0), (3, 1), (3, 2), \text{and } (3, 3)\}$. Thus, from Theorem 6.3.4, $g(3, 0) = 0$, $g(3, 1) = 2$, $g(3, 2) = 4$, and $g(3, 3) = 6$ are subgradients of θ at $\bar{u}$. Note, however, that 3/2 is also a subgradient of θ at $\bar{u}$, but 3/2 cannot be represented as $g(\bar{\mathbf{x}})$ for any $\bar{\mathbf{x}} \in X(\bar{u})$.

From the above example, it is clear that Theorem 6.3.4 gives only a sufficient characterization of subgradients. A necessary and sufficient characterization of subgradients is given in Theorem 6.3.7. First, however, the following important result is needed. The principal conclusion of this result is stated in the corollary and holds true for any arbitrary concave function θ (see Exercise 6.6). However, our proof of Theorem 6.3.6 is specialized to exploit the structure of the Lagrangian dual function θ.

6.3.6 Theorem

Let X be a nonempty compact set in R^n, and let f: $R^n \to R$ and $\boldsymbol{\beta}$: $R^n \to R^{m+\ell}$ be continuous. Let $\bar{\mathbf{w}}$, $\mathbf{d} \in R^{m+\ell}$. Then the directional derivative of θ at $\bar{\mathbf{w}}$ in the direction $\mathbf{d}$ satisfies

$$\theta'(\bar{\mathbf{w}};\mathbf{d}) \ge \mathbf{d}^t \boldsymbol{\beta}(\bar{\mathbf{x}}) \qquad \text{for some } \bar{\mathbf{x}} \in X(\bar{\mathbf{w}}).$$

Proof

Consider $\bar{\mathbf{w}} + \lambda_k \mathbf{d}$, where $\lambda_k \to 0^+$. For each k there exists an $\mathbf{x}_k \in X(\bar{\mathbf{w}} + \lambda_k \mathbf{d})$; and since X is compact, there exists a convergent subsequence $\{\mathbf{x}_k\}_{\mathscr{K}}$ having a limit $\bar{\mathbf{x}}$ in X. Given an $\mathbf{x} \in X$, note that

$$f(\mathbf{x}) + (\bar{\mathbf{w}} + \lambda_k \mathbf{d})^t \boldsymbol{\beta}(\mathbf{x}) \ge f(\mathbf{x}_k) + (\bar{\mathbf{w}} + \lambda_k \mathbf{d})^t \boldsymbol{\beta}(\mathbf{x}_k)$$

for each $k \in \mathscr{K}$. Taking the limit as $k \to \infty$, it follows that

$$f(\mathbf{x}) + \bar{\mathbf{w}}^t \boldsymbol{\beta}(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \bar{\mathbf{w}}^t \boldsymbol{\beta}(\bar{\mathbf{x}});$$

that is, $\overline{\mathbf{x}} \in X(\overline{\mathbf{w}})$. Furthermore, by the definition of $\theta(\overline{\mathbf{w}}+\lambda_k\mathbf{d})$ and $\theta(\overline{\mathbf{w}})$, we get

$$\begin{aligned}\theta(\overline{\mathbf{w}}+\lambda_k\mathbf{d})-\theta(\overline{\mathbf{w}}) &= f(\mathbf{x}_k)+(\overline{\mathbf{w}}+\lambda_k\mathbf{d})^t\beta(\mathbf{x}_k)-\theta(\overline{\mathbf{w}})\\ &\geq \lambda_k\mathbf{d}^t\beta(\mathbf{x}_k).\end{aligned}$$

The above inequality holds true for each $k \in \mathscr{K}$. Noting that $\mathbf{x}_k \to \overline{\mathbf{x}}$ as $k \in \mathscr{K}$ approaches ∞, we get

$$\lim_{\substack{k\in\mathscr{K}\\ k\to\infty}} \frac{\theta(\overline{\mathbf{w}}+\lambda_k\mathbf{d})-\theta(\overline{\mathbf{w}})}{\lambda_k} \geq \mathbf{d}^t\beta(\overline{\mathbf{x}}).$$

By Lemma 3.1.5,

$$\theta'(\overline{\mathbf{w}};\mathbf{d}) = \lim_{\lambda\to 0^+} \frac{\theta(\overline{\mathbf{w}}+\lambda\mathbf{d})-\theta(\overline{\mathbf{w}})}{\lambda}$$

exists. In view of the above inequality, the proof is complete.

Corollary

Let $\partial\theta(\overline{\mathbf{w}})$ be the collection of subgradients of θ at $\overline{\mathbf{w}}$, and suppose that the assumptions of the theorem hold true. Then

$$\theta'(\overline{\mathbf{w}};\mathbf{d}) = \inf\{\mathbf{d}^t\xi : \xi \in \partial\theta(\overline{\mathbf{w}})\}.$$

Proof

Let $\overline{\mathbf{x}}$ be as specified in the theorem. By Theorem 6.3.4, $\beta(\overline{\mathbf{x}}) \in \partial\theta(\overline{\mathbf{w}})$; and hence Theorem 6.3.6 implies that $\theta'(\overline{\mathbf{w}};\mathbf{d}) \geq \inf\{\mathbf{d}^t\xi : \xi \in \partial\theta(\overline{\mathbf{w}})\}$. Now let $\xi \in \partial\theta(\overline{\mathbf{w}})$. Since θ is concave, $\theta(\overline{\mathbf{w}}+\lambda\mathbf{d})-\theta(\overline{\mathbf{w}}) \leq \lambda\mathbf{d}^t\xi$. Dividing by $\lambda > 0$ and taking the limit as $\lambda \to 0^+$, it follows that $\theta'(\overline{\mathbf{w}};\mathbf{d}) \leq \mathbf{d}^t\xi$. Since this is true for each $\xi \in \partial\theta(\overline{\mathbf{w}})$, $\theta'(\overline{\mathbf{w}};\mathbf{d}) \leq \inf\{\mathbf{d}^t\xi : \xi \in \partial\theta(\overline{\mathbf{w}})\}$, and the proof is complete.

6.3.7 Theorem

Let X be a nonempty compact set in R^n, and let f: $R^n \to R$ and β: $R^n \to R^{m+\ell}$ be continuous. Then ξ is a subgradient of θ at $\overline{\mathbf{w}} \in R^{m+\ell}$ if and only if ξ belongs to the convex hull of $\{\beta(\mathbf{y}) : \mathbf{y} \in X(\overline{\mathbf{w}})\}$.

Proof

Denote the set $\{\beta(\mathbf{y}) : \mathbf{y} \in X(\overline{\mathbf{w}})\}$ by Λ and its convex hull by conv(Λ). By Theorem 6.3.4, $\Lambda \subseteq \partial\theta(\overline{\mathbf{w}})$; and since $\partial\theta(\overline{\mathbf{w}})$ is convex, $\text{conv}(\Lambda) \subseteq \partial\theta(\overline{\mathbf{w}})$.

Using the facts that X is compact and β is continuous, it can be verified that Λ is compact. Furthermore, the convex hull of a compact set is closed. Therefore, conv(Λ) is a closed convex set. We shall now show that $\text{conv}(\Lambda) \supseteq \partial\theta(\bar{\mathbf{w}})$.

By contradiction, suppose that there is a $\xi' \in \partial\theta(\bar{\mathbf{w}})$ but not in conv(Λ). By Theorem 2.3.4 there exist a scalar α and a nonzero vector $\mathbf{d}$ such that

$$\mathbf{d}^t\beta(\mathbf{y}) \geq \alpha \qquad \text{for each } \mathbf{y} \in X(\bar{\mathbf{w}}) \tag{6.15}$$

$$\mathbf{d}^t\xi' < \alpha. \tag{6.16}$$

By Theorem 6.3.6 there exists a $\mathbf{y} \in X(\bar{\mathbf{w}})$ such that $\theta'(\bar{\mathbf{w}};\mathbf{d}) \geq \mathbf{d}^t\beta(\mathbf{y})$; and by (6.15) we must have $\theta'(\bar{\mathbf{w}};\mathbf{d}) \geq \alpha$. But by the corollary to Theorem 6.3.6 and by (6.16), we get

$$\theta'(\bar{\mathbf{w}};\mathbf{d}) = \inf\{\mathbf{d}^t\xi : \xi \in \partial\theta(\mathbf{w})\} \leq \mathbf{d}^t\xi' < \alpha,$$

which is a contradiction. Therefore, $\xi \in \text{conv}(\Lambda)$, and $\partial\theta(\bar{\mathbf{w}}) = \text{conv}(\Lambda)$. This completes the proof.

To illustrate, consider the problem of Example 6.2.2, for which the dual function $\theta(v)$, $v \in R$, is sketched in Figure 6.4. Note that θ is differentiable (has a unique subgradient) for all v except for $v = -1$ and $v = 2$. Consider $v = 2$, for example. The set $X(2)$ is given by the set of alternative optimal solutions to the problem

$$\theta(2) = \min\{3x_2 - 6 : (x_1, x_2) \in X\}.$$

Hence, $X(2) = \{(0, 0), (4, 0)\}$, with $\theta(2) = -6$. By Theorem 6.3.4, the subgradients of the form $\beta(\bar{\mathbf{x}})$ for $\bar{\mathbf{x}} \in X(2)$ are $h(0, 0) = -3$ and $h(4, 0) = 1$. Observe that in Figure 6.4 these values are the slopes of the two affine segments defining the graph of θ that are incident at the point $(v, \theta(v)) = (2, -6)$. Therefore, as in Theorem 6.3.7, the set of subgradients of θ at $v = 2$, which is given by the slopes of the set of affine supports for the hypograph of θ, is precisely $[-3, 1]$, the set of convex combinations of -3 and 1.

For another illustration using a bivariate function θ, consider the following example.

6.3.8 Example

Consider the following primal problem:

$$\begin{aligned}
\text{Minimize } & -(x_1-4)^2-(x_2-4)^2 \\
\text{subject to } & x_1 - 3 \le 0 \\
& -x_1+x_2-2 \le 0 \\
& x_1+x_2-4 \le 0 \\
& x_1,\ x_2 \ge 0.
\end{aligned}$$

In this example, we let $g_1(x_1,x_2)=x_1-3$, $g_2(x_1,x_2)=-x_1+x_2-2$, and $X = \{(x_1,x_2): x_1+x_2-4 \le 0; x_1,x_2 \ge 0\}$. Thus, the dual function is given by

$$\theta(u_1,u_2)=\inf\{-(x_1-4)^2-(x_2-4)^2+u_1(x_1-3)+u_2(-x_1+x_2-2): \mathbf{x}\in X\}.$$

We utilize Theorem 6.3.7 to determine the set of subgradients of θ at $\bar{\mathbf{u}}=(1,5)^t$. To find the set $X(\bar{\mathbf{u}})$, we need to solve the following problem:

$$\begin{aligned}
\text{Minimize } & -(x_1-4)^2-(x_2-4)^2-4x_1+5x_2-13 \\
\text{subject to } & x_1+x_2-4 \le 0 \\
& x_1,x_2 \ge 0.
\end{aligned}$$

The objective function of this subproblem is concave, and by Theorem 3.4.7 it assumes its minimum over a compact polyhedral set at one of the extreme points. The polyhedral set X has three extreme points, (0, 0), (4, 0), and (0, 4). Noting that $f(0, 0) = f(4, 0) = -45$ and $f(0, 4) = -9$, it is evident that the optimal solutions of the above subproblem are (0, 0) and (4, 0); that is, $X(\bar{\mathbf{u}}) = \{(0, 0), (4, 0)\}$. By Theorem 6.3.7, the subgradients of θ at $\bar{\mathbf{u}}$ are thus given by the convex combinations of $\mathbf{g}(0, 0)$ and $\mathbf{g}(4, 0)$, that is, by convex combinations of the two vectors $(-3,\ -2)^t$ and $(1,\ -6)^t$. Figure 6.7 illustrates the set of subgradients.

Ascent and Steepest Ascent Directions

The dual problem is concerned with the maximization of θ subject to the constraint $\mathbf{u} \ge 0$. Given a point $\mathbf{w}^t=(\mathbf{u}^t,\mathbf{v}^t)$, we would like to investigate the directions along which θ increases. For the sake of clarity, first consider the following definition of an ascent direction, reiterated here for convenience.

6.3.9 Definition

A vector $\mathbf{d}$ is called an *ascent direction* of θ at $\mathbf{w}$ if there exists a $\delta > 0$ such that

$$\theta(\mathbf{w}+\lambda\mathbf{d}) > \theta(\mathbf{w}) \qquad \text{for each } \lambda \in (0,\ \delta).$$

Note that if θ is concave, a vector $\mathbf{d}$ is an ascent direction of θ at $\mathbf{w}$ if and only if $\theta'(\mathrm{w};\mathrm{d}) > 0$. Furthermore, θ assumes its maximum at w if and only if it has no ascent directions at w, that is, if and only if $\theta'(\mathrm{w};\mathbf{d}) \le 0$ for each $\mathbf{d}$.

Using the corollary to Theorem 6.3.6, it follows that a vector d is an ascent direction of θ at w if and only if $\inf\ \{\mathbf{d}^t\xi : \xi \in \partial\theta(\mathrm{w})\} > 0$, that is, if and only if the following inequality holds true for some $\varepsilon > 0$.

$$\mathbf{d}^t\xi \ge \varepsilon > 0 \qquad \text{for each } \xi \in \partial\theta(\mathrm{w}).$$

To illustrate, consider Example 6.3.8. The collection of subgradients of θ at the point (1, 5) is illustrated in Figure 6.7. A vector $\mathbf{d}$ is an ascent direction of θ if and only if $\mathrm{d}^t\xi \ge \varepsilon$ for each subgradient ξ, where $\varepsilon > 0$. In other words, $\mathbf{d}$ is an ascent direction if it makes an angle strictly less than 90° with each subgradient. The cone of ascent directions for this example is given in Figure 6.8. In this case, note that each subgradient is an ascent direction. However, this is not necessarily the case in general.

Since θ is to be maximized, we are interested not only in an ascent direction but also in the direction along which θ increases at the fastest local rate.

6.3.10 Definition

A vector $\bar{\mathrm{d}}$ is called a *direction of steepest ascent* of θ at w if

$$\theta'(\mathrm{w};\bar{\mathbf{d}}) = \max_{\|\mathbf{d}\|\le 1} \theta'(\mathrm{w};\mathbf{d}).$$

Theorem 6.3.11 shows that the direction of steepest ascent of the Lagrangian dual function is given by the subgradient having the smallest Euclidean norm. As evident from the proof, this result is true for any arbitrary concave function θ.

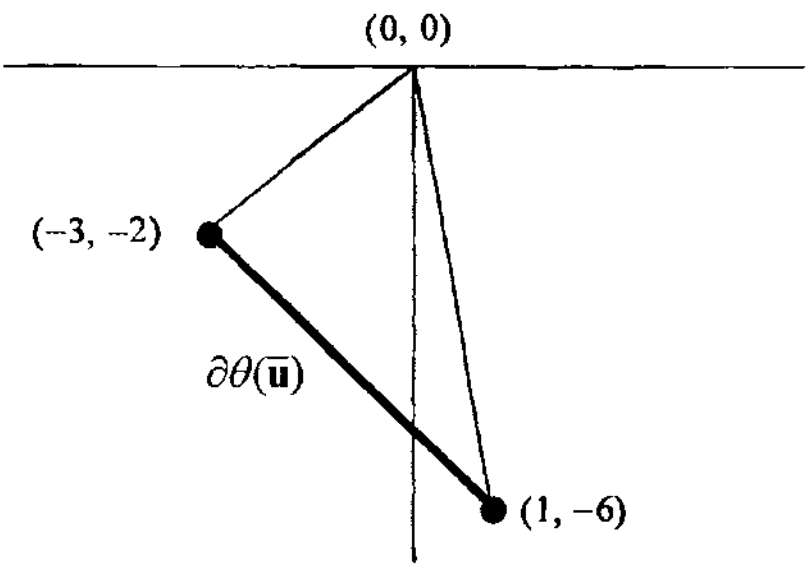


Figure 6.7 Subgradients.

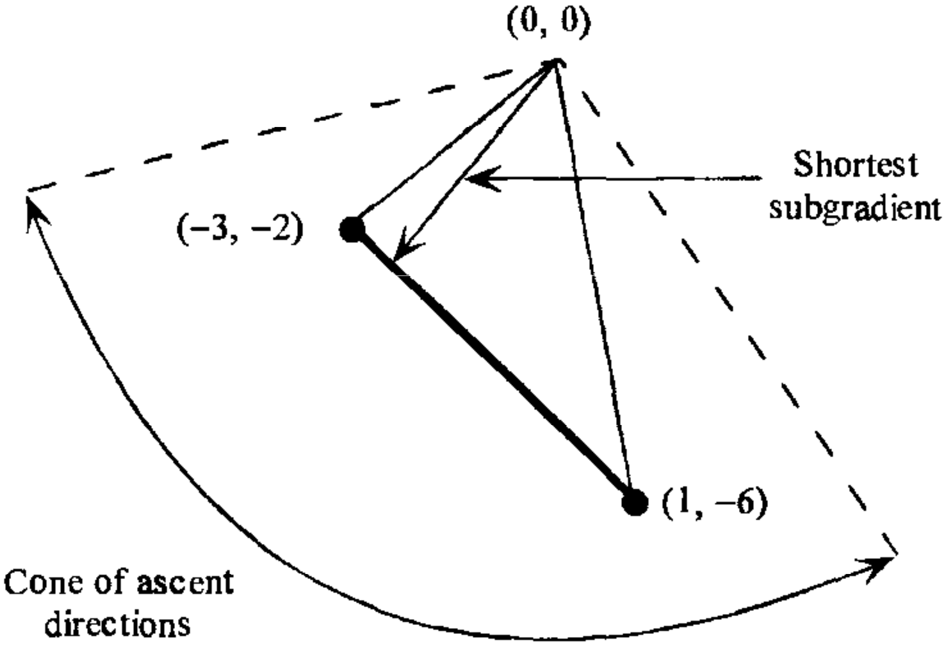


Figure 6.8 Cone of ascent directions in Example 6.3.8.

6.3.11 Theorem

Let X be a nonempty compact set in R^n, and let $f: R^n \to R$ and $\beta: R^n \to R^{m+\ell}$ be continuous. The direction of steepest ascent $\bar{\mathbf{d}}$ of θ at w is given below, where $\bar{\xi}$ is the subgradient in $\partial\theta(\mathrm{w})$ having the smallest Euclidean norm:

$$\bar{\mathbf{d}} = \begin{cases} 0 & \text{if } \bar{\xi} = 0 \\ \dfrac{\bar{\xi}}{\|\bar{\xi}\|} & \text{if } \bar{\xi} \neq \mathbf{0}. \end{cases}$$

Proof

By Definition 6.3.10 and by the corollary to Theorem 6.3.6, the steepest ascent direction can be obtained from the following expression:

$$\max_{\|\mathbf{d}\|\le 1} \theta'(\mathrm{w};\mathbf{d}) = \max_{\|\mathbf{d}\|\le 1} \inf_{\xi\in\partial\theta(\mathbf{w})} \mathbf{d}^t\xi.$$

The reader can easily verify that

$$\begin{aligned} \max_{\|\mathbf{d}\|\le 1} \theta'(\mathrm{w};\mathrm{d}) &= \max_{\|\mathbf{d}\|\le 1} \inf_{\xi\in\partial\theta(\mathbf{w})} \mathbf{d}^t\xi \\ &\le \inf_{\xi\in\partial\theta(\mathbf{w})} \max_{\|\mathbf{d}\|\le 1} \mathbf{d}^t\xi \\ &= \inf_{\xi\in\partial\theta(\mathbf{w})} \|\xi\| \\ &= \|\bar{\xi}\|. \end{aligned} \tag{6.17}$$

If we construct a direction $\bar{\mathbf{d}}$ such that $\theta'(\mathrm{w};\bar{\mathrm{d}}) = \|\bar{\xi}\|$, then by (6.17), $\bar{\mathrm{d}}$ is the steepest ascent direction. If $\bar{\xi} = \mathbf{0}$, then for $\bar{\mathbf{d}} = 0$, we obviously have $\theta'(\mathrm{w};\bar{\mathbf{d}}) = \|\bar{\xi}\|$. Now, suppose that $\bar{\xi} \neq \mathbf{0}$, and let $\bar{\mathbf{d}} = \bar{\xi}/\|\bar{\xi}\|$. Note that

$$\begin{aligned}
\theta'(\mathrm{w};\bar{\mathbf{d}}) &= \inf\{\bar{\mathbf{d}}^t\xi : \xi \in \partial\theta(\mathrm{w})\} \\
&= \inf\left\{\frac{\bar{\xi}^t\xi}{\|\bar{\xi}\|} : \xi \in \partial\theta(\mathrm{w})\right\} \\
&= \frac{1}{\|\bar{\xi}\|}\inf\left\{\|\bar{\xi}\|^2 + \bar{\xi}^t(\xi - \bar{\xi}) : \xi \in \partial\theta(\mathrm{w})\right\} \\
&= \|\bar{\xi}\| + \frac{1}{\|\bar{\xi}\|}\inf\{\bar{\xi}^t(\xi - \bar{\xi}) : \xi \in \partial\theta(\mathrm{w})\}.
\end{aligned} \tag{6.18}$$

Since $\bar{\xi}$ is the shortest vector in $\partial\theta(\mathrm{w})$, then by Theorem 2.4.1, $\bar{\xi}^t(\xi - \bar{\xi}) \geq 0$ for each $\xi \in \partial\theta(\mathrm{w})$. Hence, $\inf\{\bar{\xi}^t(\xi - \bar{\xi}) : \xi \in \partial\theta(\mathrm{w})\} = 0$ is achieved at $\bar{\xi}$. From (6.18) it then follows that $\theta'(\mathrm{w};\bar{\mathrm{d}}) = \|\bar{\xi}\|$. Thus, we have shown that the vector $\bar{\mathbf{d}}$ specified in the theorem is the direction of steepest ascent both when $\bar{\xi} = \mathbf{0}$ and when $\bar{\xi} \neq \mathbf{0}$. This completes the proof.

6.4 Formulating and Solving the Dual Problem

Given a primal Problem P to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, $\mathrm{h}(\mathbf{x}) = 0$, and $\mathbf{x} \in X$, we have defined a Lagrangian dual Problem D to maximize $\theta(\mathrm{u}, \mathrm{v})$ subject to $\mathbf{u} \geq \mathbf{0}$, where $\theta(\mathbf{u}, \mathrm{v})$ is evaluated via the (Lagrangian) subproblem $\theta(\mathrm{u}, \mathrm{v}) = \min\{f(\mathbf{x}) + \mathrm{u}^t\mathbf{g}(\mathbf{x}) + \mathrm{v}^t\mathbf{h}(\mathbf{x}) : \mathrm{x} \in X\}$. In formulating this dual problem, we have *dualized*, that is, accommodated within the Lagrangian dual objective function, the constraints $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, maintaining any other constraints within the set X. Different *formulations* of the Lagrangian dual problem might dualize different sets of constraints in constructing the Lagrangian dual function. This choice must usually be a trade-off between the ease of evaluating $\theta(\mathbf{u}, \mathrm{v})$ for a given ($\mathbf{u}$, v) versus the duality gap that might exist between P and D. For example, consider the linear discrete problem

$$\begin{aligned}
\mathrm{DP}:\quad & \text{Minimize} && \mathbf{c}^t\mathbf{x} \\
& \text{subject to} && \mathbf{A}\mathbf{x} = \mathrm{b} \\
& && \mathbf{D}\mathbf{x} = \mathrm{d} \\
& && \mathbf{x} \in X,
\end{aligned} \tag{6.19a}$$

where X is some compact, discrete set. Let us define the Lagrangian dual problem

$$\text{LDP}: \quad \text{Maximize} \quad \{\theta(\pi) : \pi \text{ unrestricted}\}, \tag{6.19b}$$

where $\theta(\pi) = \min\{\mathbf{c}^t\mathbf{x} + \pi^t(\mathbf{Ax} - \mathbf{b}) : \mathbf{Dx} = \text{d}, \mathbf{x} \in X\}$. Because of the linearity of the objective function in the latter subproblem, we equivalently have $\theta(\pi) = \min\{\mathbf{c}^t\mathbf{x} + \pi^t(\mathbf{Ax} - \mathbf{b}) : \mathbf{x} \in \text{conv}[\mathbf{x} \in X : \mathbf{Dx} = \text{d}]\}$, where conv$\{\bullet\}$ denotes the convex hull. It readily follows (see Exercise 6.7) that the Lagrangian dual objective value will match that of the modified Problem DP′ to minimize $\mathbf{c}^t\mathbf{x}$ subject to $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \in \text{conv}\{\mathbf{x} \in X : \mathbf{Dx} = \mathbf{d}\}$. Noting that DP is itself equivalent to minimizing $\mathbf{c}^t\mathbf{x}$ subject to $\mathbf{x} \in \text{conv}\{\mathbf{x} \in X: \mathbf{Ax} = \mathbf{b}, \mathbf{Dx} = \text{d}\}$, we surmise how the partial convex hull operation manifested in DP′ can influence the duality gap.

In this spirit, we may sometimes wish to manipulate the primal problem itself into a special form before constructing a Lagrangian dual formulation to create exploitable structures for the subproblem. For example, the discrete Problem DP stated above can be written equivalently as the problem to minimize $\{\mathbf{c}^t\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \text{Dy} = \text{d}, \mathbf{x} = \text{y}, \mathbf{x} \in X, \text{y} \in Y\}$, where Y is a copy of X in which the x-variables have been replaced by a set of matching y-variables. Now we can formulate a Lagrangian dual problem:

$$\overline{\text{LDP}}: \quad \text{Maximize} \quad \{\bar{\theta}(\mu) : \mu \text{ unrestricted}\}, \tag{6.20}$$

where $\bar{\theta}(\mu) \equiv \min\{\mathbf{c}^t\mathbf{x} + \mu^t(\mathbf{x} - \text{y}) : \mathbf{Ax} = \mathbf{b}, \text{Dy} = \text{d}, \mathbf{x} \in X, \text{y} \in Y\}$. Observe that this subproblem decomposes into two separable problems over the **x**- and y-variables, each with a possible specially exploitable structure. Moreover, it can be shown (see Exercise 6.8) that $\max_\mu\{\bar{\theta}(\mu)\} \ge \max_\pi \theta(\pi)$, where θ is defined in (6.19b). Hence, the Lagrangian dual formulation $\overline{\text{LDP}}$ affords a tighter representation for the primal Problem DP in the sense that it yields a smaller duality gap than does LDP. Note that, as observed previously, the value of $\overline{\text{LDP}}$ matches that of the following partial convex hull representation of the problem:

$$\text{DP}: \text{ Minimize } \{\mathbf{c}^t\mathbf{x} : \mathbf{x} \in \text{conv}\{\mathbf{x} \in X : \mathbf{Ax} = \mathbf{b}\},$$
$$\text{y} \in \text{conv}\{\text{y} \in Y : \text{Dy} = \text{d}\}, \mathbf{x} = \text{y}\}.$$

The conceptual approach leading to the formulation of $\overline{\text{LDP}}$ is called a *layering strategy* (because of the separable layers of constraints constructed), or a *Lagrangian decomposition strategy* (because of the separable decomposable structures generated). Refer to the Notes and References section for further reading on this subject matter.

Returning to the dual Problem D corresponding to the primal Problem P stated in Section 6.1, the reader will recall that we have described in the preceding section several properties of this dual function. In particular, the dual problem requires the maximization of a concave function $\theta(\mathrm{u}, \mathrm{v})$ over the simple constraint set $\{(\mathrm{u}, \mathrm{v}): \mathrm{u} \geq \mathbf{0}\}$. If θ is differentiable due to the property stated in Theorem 6.3.3, then $\nabla\theta(\overline{\mathrm{u}},\overline{\mathrm{v}})^t = [\mathbf{g}(\overline{\mathbf{x}})^t,\ \mathbf{h}(\overline{\mathbf{x}})^t]$. Various algorithms described in subsequent chapters that are applicable to maximizing differentiable concave functions can be used to solve this dual problem. These algorithms involve the generation of a suitable ascent direction d, followed by a one-dimensional line search along this direction to find a new improved solution.

To illustrate one simple scheme to find an ascent direction at a point $(\overline{\mathrm{u}}, \overline{\mathrm{v}})$, consider the following strategy. If $\nabla\theta(\overline{\mathrm{u}}, \overline{\mathrm{v}}) \neq \mathbf{0}$, then by Theorem 4.1.2 this is an ascent direction and θ will increase by moving from $(\overline{\mathrm{u}}, \overline{\mathrm{v}})$ along $\nabla\theta(\overline{\mathrm{u}}, \overline{\mathrm{v}})$. However, if some components of $\overline{\mathrm{u}}$ are equal to zero, and any of the corresponding components of $\mathbf{g}(\overline{\mathbf{x}})$ is negative, then $\overline{\mathrm{u}} + \lambda\mathbf{g}(\overline{\mathbf{x}}) \not\geq \mathbf{0}$ for $\lambda > 0$, thus violating the nonnegativity restriction. To handle this difficulty we can use a modified or projected direction $[\hat{\mathbf{g}}(\overline{\mathbf{x}}),\ \mathbf{h}(\overline{\mathbf{x}})]$, where $\hat{\mathbf{g}}(\overline{\mathbf{x}})$ is defined as

$$\hat{g}_i(\overline{\mathbf{x}}) = \begin{cases} g_i(\overline{\mathbf{x}}) & \text{if } \overline{u}_i > 0 \\ \max\{0, g_i(\overline{\mathbf{x}})\} & \text{if } \overline{u}_i = 0. \end{cases}$$

It can then be shown (see Exercise 6.9) that $[\hat{\mathbf{g}}(\overline{\mathbf{x}}),\ \mathbf{h}(\overline{\mathbf{x}})]$ is a feasible ascent direction of θ at $(\overline{\mathrm{u}}, \overline{\mathrm{v}})$. Furthermore, $[\hat{\mathbf{g}}(\overline{\mathbf{x}}),\ \mathbf{h}(\overline{\mathbf{x}})]$ is zero only when the dual maximum is reached. On the other hand, suppose that θ is nondifferentiable. In this case, the set of subgradients of θ are characterized by Theorem 6.3.7. For d to be an ascent direction of θ at (u, v), noting the corollary to Theorem 6.3.6 and the concavity of θ, we must have $\mathbf{d}^t\xi \geq \varepsilon > 0$ for each $\xi \in \partial\theta(\mathrm{u}, \mathrm{v})$. As a preliminary idea, the following problem can then be used for finding such a direction:

$$\begin{array}{lll} \text{Maximize} & \varepsilon & \\ \text{subject to} & \mathbf{d}^t\xi \geq \varepsilon & \text{for } \xi \in \partial\theta(\mathrm{u}, \mathrm{v}) \\ & d_i \geq 0 & \text{if } u_i = 0 \\ & -1 \leq d_i \leq 1 & \text{for } i = 1, \ldots, m + \ell. \end{array}$$

Note that the constraints $d_i \geq 0$ if $u_i = 0$ ensure that the vector **d** is a feasible direction, and that the normalization constraints $-1 \leq d_i \leq 1, \forall i$, guarantee a finite solution to the problem.

The reader may note the following difficulties associated with the above direction-finding problem:

1. The set $\partial\theta(\mathrm{u}, \mathrm{v})$ and hence the constraints of the problem are not known explicitly in advance. However, Theorem 6.3.7, which fully characterizes the subgradient set, could be of use.

2. The set $\partial\theta(\mathbf{u}, \mathbf{v})$ usually admits an infinite number of subgradients, so that we have a linear program having an infinite number of constraints. However, if $\partial\theta(\mathbf{u}, \mathbf{v})$ is a compact polyhedral set, then the constraints $\mathbf{d}^t\xi \geq \varepsilon$ for $\xi \in \partial\theta(\mathbf{u}, \mathbf{v})$ could be replaced by the constraints

$$\mathbf{d}^t\xi_j \geq \varepsilon \qquad \text{for } j = 1,\ldots,E,$$

where $\xi_1,\ldots,\xi_E$ are the extreme points of $\partial\theta(\mathbf{u}, \mathbf{v})$. Thus, in this case, the problem reduces to a regular linear program.

To alleviate some of the above problems, we could use a row generation strategy in which only a finite number (say, γ) of representatives of the constraint set $\mathbf{d}^t\xi \geq \varepsilon$ for $\xi \in \partial\theta(\mathbf{u}, \mathbf{v})$ are used, and the resulting direction $\mathbf{d}_\gamma$ is tested to ascertain if it is an ascent direction. This can be done by verifying if $\min\{\mathbf{d}_\gamma^t\xi : \xi \in \partial\theta(\mathbf{u}, \mathbf{v})\} > 0$. If so, $\mathbf{d}_\gamma$ can be used in the line search process. If not, the foregoing subproblem yields a subgradient $\xi_{\gamma+1}$ for which $\mathbf{d}_\gamma^t\xi_{\gamma+1} \leq 0$, and thus this constraint can be added to the direction-finding problem, and the operation could then be repeated.

We ask the reader to provide the details of this scheme in Exercise 6.30. However, this type of a procedure is fraught with computational difficulties, except for small problems having simple structures. In Chapter 8 we address more sophisticated and efficient subgradient-based optimization schemes that can be used to optimize θ whenever it is nondifferentiable. These procedures employ various strategies for constructing directions based on single subgradients, possibly deflected by suitable means, or based on a bundle of subgradients collected over some local neighborhood. The directions need not always be ascent directions, but ultimate convergence to an optimum is nevertheless assured. We refer the reader to Chapter 8 and its Notes and References section for further information on this topic.

We now proceed to describe in detail one particular cutting plane or outer-linearization scheme for solving the dual Problem D. The concept of this approach is important in its own right, as it constitutes a useful ingredient for many decomposition and partitioning methods.

Cutting Plane or Outer-Linearization Method

The methods discussed in principle above for solving the dual problem generate at each iteration a direction of motion and adopt a step size along this direction, with a view ultimately to finding the maximum for the Lagrangian dual function. We now discuss a strategy for solving the dual problem in which, at each iteration, a function that approximates the dual function is optimized.

Recall that the dual function θ is defined by

$$\theta(\mathbf{u}, \mathbf{v}) = \inf\{f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}.$$

Letting $z = \theta(\mathbf{u}, \mathbf{v})$, the inequality $z \le f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$ must hold true for each $\mathbf{x} \in X$. Hence, the dual problem of maximizing $\theta(\mathbf{u}, \mathbf{v})$ over $\mathbf{u} \ge \mathbf{0}$ is equivalent to the following problem:

$$\begin{aligned} &\text{Maximize} && z \\ &\text{subject to} && z \le f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \quad \text{for } \mathbf{x} \in X \\ &&& \mathbf{u} \ge \mathbf{0}. \end{aligned} \tag{6.21}$$

Note that the above problem is a linear program in the variables z, $\mathbf{u}$, and $\mathbf{v}$. Unfortunately, however, the constraints are infinite in number and are not known explicitly. Suppose that we have the points $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$ in X, and consider the following approximating problem:

$$\begin{aligned} &\text{Maximize} && z \\ &\text{subject to} && z \le f(\mathbf{x}_j) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_j) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_j) \quad \text{for } j = 1, \ldots, k-1 \\ &&& \mathbf{u} \ge \mathbf{0}. \end{aligned} \tag{6.22}$$

The above problem is a linear program having a finite number of constraints and can be solved by the simplex method, for example. Let $(z_k, \mathbf{u}_k, \mathbf{v}_k)$ be an optimal solution to this approximating problem, sometimes referred to as the *master program*. If this solution satisfies (6.21), then it is an optimal solution to the Lagrangian dual problem. To check whether (6.21) is satisfied, consider the following *subproblem*:

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) + \mathbf{u}_k^t\mathbf{g}(\mathbf{x}) + \mathbf{v}_k^t\mathbf{h}(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in X. \end{aligned}$$

Let $\mathbf{x}_k$ be an optimal solution to the above problem, so that $\theta(\mathbf{u}_k, \mathbf{v}_k) = f(\mathbf{x}_k) + \mathbf{u}_k^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k^t\mathbf{h}(\mathbf{x}_k)$. If $z_k \le (\mathbf{u}_k, \mathbf{v}_k)$, then $(\mathbf{u}_k, \mathbf{v}_k)$ is an optimal solution to the Lagrangian dual problem. Otherwise, for $(\mathbf{u}, \mathbf{v}) = (\mathbf{u}_k, \mathbf{v}_k)$, the inequality (6.21) is not satisfied for $\mathbf{x} = \mathbf{x}_k$. Thus, we add the constraint

$$z \le f(\mathbf{x}_k) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_k)$$

to the constraints in (6.22), and re-solve the master linear program. Obviously, the current optimal point $(z_k, \mathbf{u}_k, \mathbf{v}_k)$ contradicts this added constraint. Thus, this point is cut away, hence the name *cutting plane algorithm*.

Summary of the Cutting Plane or Outer-Linearization Method

Assume that f, $\mathbf{g}$, and $\mathbf{h}$ are continuous and that X is compact, so that the set $X(\mathbf{u}, \mathbf{v})$ is nonempty for each $(\mathbf{u}, \mathbf{v})$.

Initialization Step Find a point $\mathbf{x}_0 \in X$ such that $\mathbf{g}(\mathbf{x}_0) \le \mathbf{0}$ and $\mathbf{h}(\mathbf{x}_0) = \mathbf{0}$. Let $k = 1$, and go to the main step.

Main Step Solve the following *master program*:

$$\begin{aligned} &\text{Maximize} && z \\ &\text{subject to} && z \le f(\mathbf{x}_j) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_j) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_j) \qquad \text{for } j = 0,\dots,k-1 \\ &&& \mathbf{u} \ge \mathbf{0}. \end{aligned}$$

Let $(z_k, \mathbf{u}_k, \mathbf{v}_k)$ be an optimal solution. Solve the following *subproblem*:

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) + \mathbf{u}_k^t\mathbf{g}(\mathbf{x}) + \mathbf{v}_k^t\mathbf{h}(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in X. \end{aligned}$$

Let $\mathbf{x}_k$ be an optimal point, and let $\theta(\mathbf{u}_k, \mathbf{v}_k) = f(\mathbf{x}_k) + \mathbf{u}_k^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k^t\mathbf{h}(\mathbf{x}_k)$. If $z_k = \theta(\mathbf{u}_k, \mathbf{v}_k)$, then stop; $(\mathbf{u}_k, \mathbf{v}_k)$ is an optimal dual solution. Otherwise, if $z_k > \theta(\mathbf{u}_k, \mathbf{v}_k)$, then add the constraint $z \le f(\mathbf{x}_k) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_k)$ to the master program, replace k by $k + 1$, and repeat the main step.

At each iteration, a cut (constraint) is added to the master problem, and hence the size of the master problem increases monotonically. In practice, if the size of the master problem becomes excessively large, all constraints that are not binding may be thrown away. Theoretically, this might not guarantee convergence, unless, for example, the dual value has strictly increased since the last time such a deletion was executed, and the set X has a finite number of elements. (See Exercise 6.28; and for a general convergence theorem, see Exercises 7.21 and 7.22.) Also, note that the optimal solution values of the master problem form a nonincreasing sequence $\{z_k\}$. Since each z_k is an upper bound on the optimal value of the dual problem, we may stop after iteration k if $z_k - \max_{1 \le j \le k} \theta(\mathbf{u}_j, \mathbf{v}_j) < \varepsilon$, where ε is a small positive number .

Interpretation as a Tangential Approximation or Outer-Linearization Technique

The foregoing algorithm for maximizing the dual function can be interpreted as a tangential approximation technique. By the definition of θ, we must have

$$\theta(\mathbf{u}, \mathbf{v}) \le f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \qquad \text{for } \mathbf{x} \in X.$$

Thus, for any fixed $\mathbf{x} \in X$, the hyperplane

$$\{(\mathbf{u}, \mathbf{v}, z) : \mathbf{u} \in R^m,\ \mathbf{v} \in R^\ell,\ z = f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})\}$$

bounds the function θ from above.

The master problem at iteration k is equivalent to solving the following problem:

$$\begin{aligned} &\text{Maximize} \quad \hat{\theta}(\mathbf{u}, \mathbf{v}) \\ &\text{subject to} \quad \mathbf{u} \geq \mathbf{0}, \end{aligned}$$

where $\hat{\theta}(\mathbf{u}, \mathbf{v}) = \min\{f(\mathbf{x}_j) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_j) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_j) : j = 1, \ldots, k-1\}$. Note that $\hat{\theta}$ is a piecewise linear function that provides an *outer approximation* or *outer linearization* for θ by considering only $k-1$ of the bounding hyperplanes.

Let the optimal solution to the master problem be $(z_k, \mathbf{u}_k, \mathbf{v}_k)$. Now the subproblem is solved yielding $\theta(\mathbf{u}_k, \mathbf{v}_k)$ and $\mathbf{x}_k$. If $z_k > \theta(\mathbf{u}_k, \mathbf{v}_k)$, the new constraint $z \leq f(\mathbf{x}_k) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_k)$ is added to the master problem, giving a new and tighter piecewise linear approximation to θ. Since $\theta(\mathbf{u}_k, \mathbf{v}_k) = f(\mathbf{x}_k) + \mathbf{u}_k^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k^t\mathbf{h}(\mathbf{x}_k)$, the hyperplane $\{(z, \mathbf{u}, \mathbf{v}) : z = f(\mathbf{x}_k) + \mathbf{u}^t\mathbf{g}(\mathbf{x}_k) + \mathbf{v}^t\mathbf{h}(\mathbf{x}_k)\}$ is tangential to the graph of θ at $(z_k, \mathbf{u}_k, \mathbf{v}_k)$: hence, the name *tangential approximation.*

6.4.1 Example

$$\begin{aligned} &\text{Minimize} \quad (x_1-2)^2 + (1/4)x_2^2 \\ &\text{subject to} \quad x_1 - (7/2)x_2 - 1 \leq 0 \\ &\qquad\qquad\quad 2x_1 + 3x_2 = 4. \end{aligned}$$

We let $X = \{(x_1, x_2) : 2x_1 + 3x_2 = 4\}$, so that the Lagrangian dual function is given by

$$\theta(u) = \min\{(x_1-2)^2 + (1/4)x_2^2 + u(x_1 - (7/2)x_2 - 1) : 2x_1 + 3x_2 = 4\}. \quad (6.23)$$

The cutting plane method is initialized with a feasible solution $\mathbf{x}_0 = (5/4, 1/2)^t$. At Step 1 of the first iteration, we solve the following problem:

$$\begin{aligned} &\text{Maximize} \quad z \\ &\text{subject to} \quad z \leq 5/8 - (3/2)u \\ &\qquad\qquad\quad u \geq 0. \end{aligned}$$

The optimal solution is $(z_1, u_1) = (5/8, 0)$. At Step 2 we solve (6.23) for $u = u_1 = \mathbf{0}$, yielding an optimal solution $\mathbf{x}_1 = (2, 0)^t$ with $\theta(u_1) = 0 < z_1$. Hence, more iterations are needed. A summary of the first four iterations is given in Table 6.1.

The approximating function $\hat{\theta}$ at the end of the fourth iteration is shown by darkened lines in Figure 6.9. The reader can easily verify that the Langrangian dual function for this problem is given by $\theta(u) = -(5/2)u^2 + u$ and that the hyperplanes added at Iteration 2 onward are indeed tangential to the graph of θ

Table 6.1 Summary of Computations for Example 6.4.1

		Step 1 Solution		Step 2 Solution
Iteration k	Constraint Added	(z_k, u_k)	$\mathbf{x}_k^t$	$\theta(u_k)$
1	$z \le 5/8-(3/2)u$	(5/8, 0)	(2, 0)	0
2	$z \le 0+u$	(1/4, 1/4)	(13/8, 1/4)	3/32
3	$z \le 5/32-(1/4)u$	(1/8, 1/8)	(29/16, 1/8)	11/128
4	$z \le 5/128+(3/8)u$	(7/64, 3/16)	(55/32, 3/16)	51/512

at the respective points (z_k, u_k). Incidentally, the dual objective function is maximized at $\bar{u} = 1/5$ with $\theta(\bar{u}) = 1/10$. Note that the sequence $\{u_k\}$ converges to the optimal point $\bar{u} = 1/5$.

6.5 Getting the Primal Solution

Thus far we have studied several properties of the dual function and described some procedures for solving the dual problem. However, our main concern is finding an optimal solution to the primal problem.

In this section we develop some theorems that will aid us in finding a solution to the primal problem as well as solutions to perturbations of the primal problem. However, for nonconvex programs, as a result of the possible presence of a duality gap, additional work is usually needed to find an optimal primal solution.

Solutions to Perturbed Primal Problems

During the course of solving the dual problem, the following problem, which is used to evaluate the function θ at (u, v), is solved frequently:

$$\begin{aligned} &\text{Minimize} \quad f(\mathbf{x})+\mathbf{u}^t\mathbf{g}(\mathbf{x})+\mathbf{v}^t\mathbf{h}(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{x} \in X. \end{aligned}$$

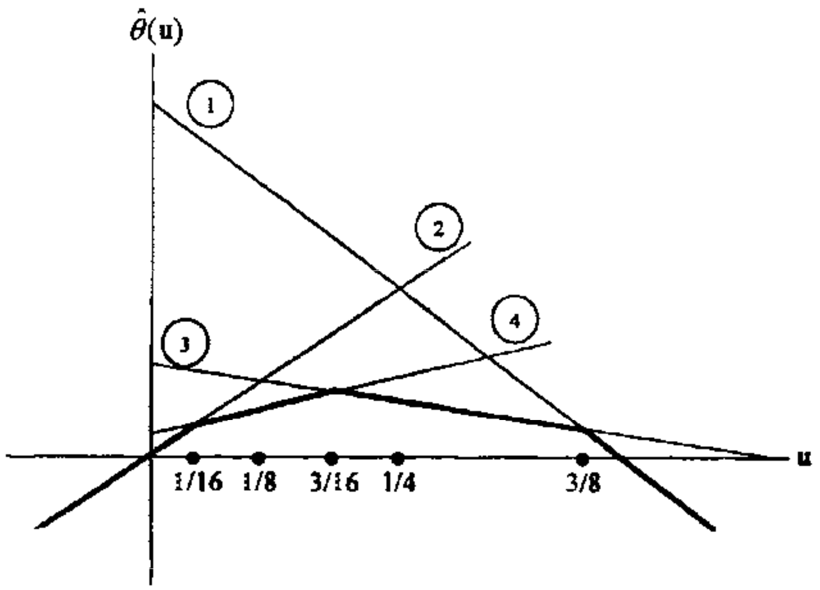


Figure 6.9 Tangential approximation of θ.

Theorem 6.5.1 shows that an optimal solution $\bar{\mathbf{x}}$ to the above problem is also an optimal solution to a problem that is similar to the primal problem, in which some of the constraints are perturbed. Specifically, $\bar{\mathbf{x}}$ evaluates $v[\mathbf{g}(\bar{\mathbf{x}}), \mathbf{h}(\bar{\mathbf{x}})]$, where v is the perturbation function defined in (6.9).

6.5.1 Theorem

Let (u, v) be a given vector with $\mathbf{u} \geq \mathbf{0}$. Consider the problem to minimize $f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$ subject to $\mathbf{x} \in X$. Let $\bar{\mathbf{x}}$ be an optimal solution. Then $\bar{\mathbf{x}}$ is an optimal solution to the following problem, where $I = \{i : u_i > 0\}$:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \leq g_i(\bar{\mathbf{x}}) & \text{for } i \in I \\ & h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}}) & \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. & \end{array}$$

In particular, $\bar{\mathbf{x}}$ solves the problem to evaluate $v[\mathbf{g}(\bar{\mathbf{x}}), \mathbf{h}(\bar{\mathbf{x}})]$, where v is the perturbation function defined in (6.9).

Proof

Let $\mathbf{x} \in X$ be such that $h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}})$ for $i = 1,\ldots, \ell$ and $g_i(\mathbf{x}) \leq g_i(\bar{\mathbf{x}})$ for $i \in I$. Note that

$$f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{v}^t\mathbf{h}(\bar{\mathbf{x}}). \tag{6.24}$$

But since $\mathbf{h}(\mathbf{x}) = \mathbf{h}(\bar{\mathbf{x}})$ and $\mathbf{u}^t\mathbf{g}(\mathbf{x}) = \sum_{i \in I} u_i g_i(\mathbf{x}) \leq \sum_{i \in I} u_i g_i(\bar{\mathbf{x}}) = \mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}})$, we get from (6.24) that

$$f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}) \geq f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}),$$

which shows that $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$. Hence, $\bar{\mathbf{x}}$ solves the problem stated in the theorem. Moreover, since this problem is a relaxation of (6.9) for $\mathbf{y} = \bar{\mathbf{y}}$, where $\bar{\mathbf{y}}^t = [\mathbf{g}(\bar{\mathbf{x}})^t, \mathbf{h}(\bar{\mathbf{x}})^t]$, and since $\bar{\mathbf{x}}$ is feasible to (6.9) with $\mathbf{y} = \bar{\mathbf{y}}$, it follows that $\bar{\mathbf{x}}$ evaluates $v(\bar{\mathbf{y}})$. This completes the proof.

Corollary

Under the assumptions of the theorem, suppose that $\mathbf{g}(\bar{\mathbf{x}}) \leq \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{u}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$. Then $\bar{\mathbf{x}}$ is an optimal solution to the following problem:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & \text{for } i \in I \\ & h_i(\mathbf{x}) = 0 & \text{for } i = 1,\ldots,\ell \\ & \mathbf{x} \in X. & \end{array}$$

In particular, $\overline{\mathbf{x}}$ is an optimal solution to the original primal problem, and (u, v) is an optimal solution to the dual problem.

Proof

Note that $\mathbf{u}^t\mathbf{g}(\overline{\mathbf{x}}) = 0$ implies that $g_i(\overline{\mathbf{x}}) = 0$ for $i \in I$; and from the theorem, it follows that $\overline{\mathbf{x}}$ solves the problem stated. Also, since the feasible region of the primal problem is contained in that of the above problem, and since $\overline{\mathbf{x}}$ is a feasible solution to the primal problem, then $\overline{\mathbf{x}}$ is an optimal solution to the primal problem. Furthermore, $f(\overline{\mathbf{x}}) = f(\overline{\mathbf{x}}) + \mathbf{u}^t\mathbf{g}(\overline{\mathbf{x}}) + \mathbf{v}^t\mathbf{h}(\overline{\mathbf{x}}) = \theta(\mathbf{u},\mathbf{v})$, so that (u, v) solves the dual problem. This completes the proof.

Of course, the conditions of the above corollary coincide precisely with the saddle point optimality conditions (a), (b), and (c) of Theorem 6.2.5, implying that $(\overline{\mathbf{x}}, \mathbf{u}, \mathbf{v})$ is a saddle point and, hence, that $\overline{\mathbf{x}}$ and (u, v) solve Problems P and D, respectively. Also, elements of the proof of Theorem 6.5.1 are evident in the proof of Theorem 6.2.7. However, our purpose in highlighting Theorem 6.5.1 and its corollary is to emphasize the role played by this result in deriving heuristic primal solutions based on solving the dual problem. As seen from Theorem 6.5.1, as the dual function θ is evaluated at a given point (u, v), we obtain a point $\overline{\mathbf{x}}$ that is an optimal solution to a problem that is closely related to the original problem, in which the constraints are perturbed from $\mathbf{h}(\mathbf{x}) = 0$ and $g_i(\mathbf{x}) \le 0$ for $i = 1,\ldots, m$, to $\mathbf{h}(\mathbf{x}) = \mathbf{h}(\overline{\mathbf{x}})$ and $g_i(\mathbf{x}) \le g_i(\overline{\mathbf{x}})$ for $i = 1,\ldots, m$.

In particular, during the course of solving the dual problem, suppose that for a given (u, v) with $\mathbf{u} \ge \mathbf{0}$, we have $\hat{\mathbf{x}} \in X(\mathbf{u}, \mathbf{v})$. Furthermore, for some $\varepsilon > 0$, suppose that $|g_i(\hat{\mathbf{x}})| \le \varepsilon$ for $i \in I$, $g_i(\hat{\mathbf{x}}) \le \varepsilon$ for $i \notin I$, and $|h_i(\hat{\mathbf{x}})| \le \varepsilon$ for $i = 1,\ldots, \ell$. Note that if ε is sufficiently small, then $\hat{\mathbf{x}}$ is *near-feasible*. Now, suppose that $\overline{\mathbf{x}}$ is an optimal solution to the primal Problem P. Then, by the definition of $\theta(\mathbf{u}, \mathbf{v})$,

$$f(\hat{\mathbf{x}}) + \sum_{i\in I} u_i g_i(\hat{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i h_i(\hat{\mathbf{x}}) \le f(\overline{\mathbf{x}}) + \sum_{i\in I} u_i g_i(\overline{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i h_i(\overline{\mathbf{x}}) \le f(\overline{\mathbf{x}})$$

since $h_i(\overline{\mathbf{x}}) = 0$, $g_i(\overline{\mathbf{x}}) \le 0$, and $u_i \ge 0$. The above inequality thus implies that

$$f(\hat{\mathbf{x}}) \le f(\overline{\mathbf{x}}) + \varepsilon\left[\sum_{i\in I} u_i + \sum_{i=1}^{\ell} |v_i|\right].$$

Therefore, if ε is sufficiently small so that $\varepsilon[\sum_{i\in I} u_i + \sum_{i=1}^{\ell}|v_i|]$ is small enough, then $\hat{\mathbf{x}}$ is a *near-optimal solution*. In many practical problems, such a solution is often acceptable.

Note also that in the absence of a duality gap, if $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ are, respectively, optimal primal and dual solutions, then, by Theorem 6.2.5, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ is a saddle point. Hence, by Property (a) of Theorem 6.2.5, $\bar{\mathbf{x}}$ minimizes $\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ over $\mathbf{x} \in X$. This means that there exists an optimal solution to the primal problem among points in the set $X(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, where $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ is an optimal solution to the dual problem. Of course, not any solution $\bar{\mathbf{x}} \in X(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solves the primal problem unless $\bar{\mathbf{x}}$ is feasible to P and it satisfies the complementary slackness condition $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$.

Generating Primal Feasible Solutions in the Convex Case

The foregoing discussion was concerned with general, perhaps nonconvex problems. Under suitable convexity assumptions, we can easily obtain primal feasible solutions at each iteration of the dual problem by solving a linear program. In particular, suppose that we are given a point $\mathbf{x}_0$, which is feasible to the original problem, and let the points $\mathbf{x}_j \in X(\mathbf{u}_j, \mathbf{v}_j)$ for $j = 1,\ldots, k$ be generated by an arbitrary algorithm used to maximize the dual function. Theorem 6.5.2 shows that a feasible solution to the primal problem can be obtained by solving the following linear programming problem P':

$$\begin{aligned}
\text{P}': \quad \text{Minimize} \quad & \sum_{j=0}^{k} \lambda_j f(\mathbf{x}_j) \\
\text{subject to} \quad & \sum_{j=0}^{k} \lambda_j \mathbf{g}(\mathbf{x}_j) \le \mathbf{0} \\
& \sum_{j=0}^{k} \lambda_j \mathbf{h}(\mathbf{x}_j) = \mathbf{0} \\
& \sum_{j=0}^{k} \lambda_j = 1 \\
& \lambda_j \ge 0 \qquad \text{for } j = 0,\ldots,k.
\end{aligned} \tag{6.25}$$

6.5.2 Theorem

Let X be a nonempty convex set in R^n, let f: $R^n \to R$ and $\mathbf{g}$: $R^n \to R^m$ be convex, and let $\mathbf{h}$: $R^n \to R^{\ell}$ be affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Let $\mathbf{x}_0$ be an initial feasible solution to Problem P, and suppose that $\mathbf{x}_j \in X(\mathbf{u}_j, \mathbf{v}_j)$ for $j = 1,\ldots, k$ are generated by any algorithm for solving the dual problem.

Furthermore, let $\bar{\lambda}_j$ for $j = 0,\ldots, k$ be an optimal solution to Problem P′ defined in (6.25), and let $\bar{\mathbf{x}}_k = \sum_{j=0}^{k} \bar{\lambda}_j \mathbf{x}_j$. Then $\bar{\mathbf{x}}_k$ is a feasible solution to the primal Problem P. Furthermore, letting $z_k = \sum_{j=0}^{k} \bar{\lambda}_j f(\mathbf{x}_j)$ and $z^* = \inf\{f(\mathbf{x}) : \mathbf{x} \in X,$ $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, if $z_k - \theta(\mathbf{u}, \mathbf{v}) \leq \varepsilon$ for some (u, v) with $\mathbf{u} \geq \mathbf{0}$, then $f(\bar{\mathbf{x}}_k) \leq z^* + \varepsilon$.

Proof

Since X is convex and $\mathbf{x}_j \in X$ for each j, $\bar{\mathbf{x}}_k \in X$. Since g is convex and h is affine, and noting the constraints of Problem P′, $\mathbf{g}(\bar{\mathbf{x}}_k) \leq \mathbf{0}$ and $\mathbf{h}(\bar{\mathbf{x}}_k) = \mathbf{0}$. Thus, $\bar{\mathbf{x}}_k$ is a feasible solution to the primal problem. Now suppose that $z_k - \theta(\mathbf{u}, \mathbf{v}) \leq \varepsilon$ for some (u, v) with $\mathbf{u} \geq \mathbf{0}$. Noting the convexity of f and Theorem 6.2.1, we get

$$f(\bar{\mathbf{x}}_k) \leq \sum_{j=0}^{k} \bar{\lambda}_j f(\mathbf{x}_j) = z_k \leq \theta(\mathbf{u}, \mathbf{v}) + \varepsilon \leq z^* + \varepsilon$$

and the proof is complete.

At each iteration of the dual maximization problem, we can thus obtain a primal feasible solution by solving the linear programming Problem P′. Even though the primal objective values $\{f(\bar{\mathbf{x}}_k)\}$ of the generated primal feasible points are not necessarily decreasing, they form a sequence that is bounded from above by the nonincreasing sequence $\{z_k\}$.

Note that if z_k is close enough to the dual objective value evaluated at any dual feasible point (u, v), where $\mathbf{u} \geq \mathbf{0}$, then $\bar{\mathbf{x}}_k$ is a near-optimal primal feasible solution. Also note that we need not solve Problem P′ in the case of the cutting plane algorithm, since it is precisely the linear programming dual of the master problem stated in Step 1 of this algorithm. Thus, the optimal variables $\bar{\lambda}_0,\ldots,\bar{\lambda}_k$ can be retrieved easily from the solution to the master problem, and $\bar{\mathbf{x}}_k$ can be computed as $\sum_{j=0}^{k} \bar{\lambda}_j \mathbf{x}_j$. It is also worth mentioning that the termination criterion $z_k = \theta(\mathbf{u}_k, \mathbf{v}_k)$ in the cutting plane algorithm can be interpreted as letting $(\mathbf{u}, \mathbf{v}) = (\mathbf{u}_k, \mathbf{v}_k)$ and $\varepsilon = 0$ in the Theorem 6.5.2.

To illustrate the above procedure, consider Example 6.4.1. At the end of Iteration $k = 1$, we have the points $\mathbf{x}_0 = (5/4, 1/2)^t$ and $\mathbf{x}_1 = (2, 0)^t$. The associated primal point $\bar{\mathbf{x}}_1$ can be obtained by solving the following linear programming problem:

$$\begin{aligned} \text{Minimize} \quad & (5/8)\lambda_0 \\ \text{subject to} \quad & -(3/2)\lambda_0 + \lambda_1 \le 0 \\ & \lambda_0 + \lambda_1 = 1 \\ & \lambda_0, \lambda_1 \ge 0. \end{aligned}$$

The optimal solution to this problem is given by $\bar{\lambda}_0 = 2/5$ and $\bar{\lambda}_1 = 3/5$. This yields a primal feasible solution

$$\bar{\mathbf{x}}_1 = (2/5)(5/4, 1/2)^t + (3/5)(2, 0)^t = (17/10, 2/10)^t.$$

As pointed out earlier, the above linear program need not be solved separately to find the values of $\bar{\lambda}_0$ and $\bar{\lambda}_1$, since its dual has already been solved during the course of the cutting plane algorithm.

6.6 Linear and Quadratic Programs

In this section, we discuss some special cases of Lagrangian duality. In particular, we discuss briefly duality in linear and quadratic programming. For linear programming problems, we relate the Lagrangian dual to that derived in Chapter 2 (see Theorem 2.7.3 and its corollaries). In the case of quadratic programming problems, we derive the well-known Dorn's dual program via Lagrangian duality.

Linear Programming

Consider the following primal linear program:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^t\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \ge \mathbf{0}. \end{aligned}$$

Letting $X = \{\mathbf{x} : \mathbf{x} \ge \mathbf{0}\}$, the Lagrangian dual of this problem is to maximize $\theta(\mathbf{v})$, where

$$\theta(\mathbf{v}) = \inf\{\mathbf{c}^t\mathbf{x} + \mathbf{v}^t(\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \ge \mathbf{0}\} = \mathbf{v}^t\mathbf{b} + \inf\{(\mathbf{c}^t - \mathbf{v}^t\mathbf{A})\mathbf{x} : \mathbf{x} \ge \mathbf{0}\}.$$

Clearly,

$$\theta(\mathbf{v}) = \begin{cases} \mathbf{v}^t\mathbf{b} & \text{if } (\mathbf{c}^t - \mathbf{v}^t\mathbf{A}) \ge \mathbf{0} \\ -\infty & \text{otherwise.} \end{cases}$$

Hence, the dual problem can be stated as follows:

$$\begin{aligned} \text{Maximize} \quad & \mathbf{v}^t\mathbf{b} \\ \text{subject to} \quad & \mathbf{A}^t\mathbf{v} \le \mathbf{c}. \end{aligned}$$

Recall that this is precisely the *dual problem* discussed in Section 2.7. Thus, in the case of linear programs, the dual problem does not involve the primal variables. Furthermore, the dual problem itself is a linear program, and the reader can verify that the dual of the dual problem is the original primal program. Theorem 6.6.1 summarizes the relationships between the primal and dual problems as established by Theorem 2.7.3 and its three corollaries.

6.6.1 Theorem

Consider the primal and dual linear problems stated above. One of the following mutually exclusive cases will occur:

1. The primal problem admits a feasible solution and has an unbounded objective value, in which case the dual problem is infeasible.
2. The dual problem admits a feasible solution and has an unbounded objective value, in which case the primal problem is infeasible.
3. Both problems admit feasible solutions, in which case both problems have optimal solutions $\overline{\mathbf{x}}$ and $\overline{\mathbf{v}}$ such that $\mathbf{c}^t\overline{\mathbf{x}} = \mathbf{b}^t\overline{\mathbf{v}}$ and $(\mathbf{c}^t - \mathbf{v}^t\mathbf{A})\overline{\mathbf{x}} = 0$.
4. Both problems are infeasible.

Proof

See Theorem 2.7.3 and its Corollaries 1 and 3.

Quadratic Programming

Consider the following quadratic programming problem:

$$\begin{aligned} &\text{Minimize} && (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x} + \mathbf{d}^t\mathbf{x} \\ &\text{subject to} && \mathbf{A}\mathbf{x} \le \mathbf{b}, \end{aligned}$$

where $\mathbf{H}$ is symmetric and positive semidefinite, so that the objective function is convex. The Lagrangian dual problem is to maximize $\theta(\mathbf{u})$ over $\mathbf{u} \ge \mathbf{0}$, where

$$\theta(\mathbf{u}) = \inf\{(1/2)\mathbf{x}^t\mathbf{H}\mathbf{x} + \mathbf{d}^t\mathbf{x} + \mathbf{u}^t(\mathbf{A}\mathbf{x} - \mathbf{b}) : \mathbf{x} \in R^n\}. \tag{6.26}$$

Note that for a given $\mathbf{u}$, the function $(1/2)\mathbf{x}^t\mathbf{H}\mathbf{x} + \mathbf{d}^t\mathbf{x} + \mathbf{u}^t(\mathbf{A}\mathbf{x} - \mathbf{b})$ is convex, so a necessary and sufficient condition for a minimum is that the gradient must vanish; that is,

$$\mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} + \mathbf{d} = \mathbf{0}. \tag{6.27}$$

Thus, the dual problem can be written as follows:

$$\begin{aligned} \text{Maximize} \quad & (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x} + \mathbf{d}^t\mathbf{x} + \mathbf{u}^t(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ \text{subject to} \quad & \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} = -\mathbf{d} \\ & \mathbf{u} \geq \mathbf{0}. \end{aligned} \tag{6.28}$$

Now, from (6.27), we have $\mathbf{d}^t\mathbf{x} + \mathbf{u}^t\mathbf{A}\mathbf{x} = -\mathbf{x}^t\mathbf{H}\mathbf{x}$. Substituting this into (6.28), we derive the familiar form of *Dorn's dual quadratic program*:

$$\begin{aligned} \text{Maximize} \quad & -(1/2)\mathbf{x}^t\mathbf{H}\mathbf{x} - \mathbf{b}^t\mathbf{u} \\ \text{subject to} \quad & \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} = -\mathbf{d} \\ & \mathbf{u} \geq \mathbf{0}. \end{aligned} \tag{6.29}$$

Again, by Lagrangian duality, if one problem is unbounded, then the other is infeasible. Moreover, following Theorem 6.2.6, if both problems are feasible, then they both have optimal solutions having the same objective value.

We now develop an alternative form of the Lagrangian dual problem under the assumption that $\mathbf{H}$ is positive definite, so that $\mathbf{H}^{-1}$ exists. In this case, the unique solution to (6.27) is given by

$$\mathbf{x} = -\mathbf{H}^{-1}(\mathbf{d} + \mathbf{A}^t\mathbf{u}).$$

Substituting in (6.26), it follows that

$$\theta(\mathbf{u}) = (1/2)\mathbf{u}^t\mathbf{D}\mathbf{u} + \mathbf{u}^t\mathbf{c} - (1/2)\mathbf{d}^t\mathbf{H}^{-1}\mathbf{d},$$

where $\mathbf{D} = -\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^t$ and $\mathbf{c} = -\mathbf{b} - \mathbf{A}\mathbf{H}^{-1}\mathbf{d}$. The dual problem is thus given by:

$$\begin{aligned} \text{Maximize} \quad & (1/2)\mathbf{u}^t\mathbf{D}\mathbf{u} + \mathbf{u}^t\mathbf{c} - (1/2)\mathbf{d}^t\mathbf{H}^{-1}\mathbf{d} \\ \text{subject to} \quad & \mathbf{u} \geq \mathbf{0}. \end{aligned} \tag{6.30}$$

The dual problem (6.30) can be solved relatively easily using the algorithms described in Chapters 8 through 11, noting that this problem simply seeks to maximize a concave quadratic function over the nonnegative orthant. (See Exercise 6.45 for a simplified scheme.)

Exercises

[6.1] Consider the (singly) constrained problem to minimize $f(\mathbf{x})$ subject to $g(\mathbf{x}) \leq 0$ and $\mathbf{x} \in X$. Define $G = \{(y, z) : y = g(\mathbf{x}), z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in X\}$, and let $v(y) = \min\{f(\mathbf{x}) : g(\mathbf{x}) \leq y, \mathbf{x} \in X\}$, $y \in R$, be the associated perturbation function. Show that v is the pointwise supremum over all possible nonincreasing functions whose epigraph contains G.

[6.2] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_1(\mathbf{x}) \leq 0$ and $g_2(\mathbf{x}) \leq 0$ as illustrated in Figure 4.13. Denote $X = \{\mathbf{x} : g_1(\mathbf{x}) \leq 0\}$. Sketch the perturbation function $v(y) = \min\{f(\mathbf{x}) : g_2(\mathbf{x}) \leq y, \mathbf{x} \in X\}$ and indicate the duality gap. Provide a possible sketch for the set $G = \{(y, z) : y = g_2(\mathbf{x}),\ z = f(\mathbf{x})$ for some $\mathbf{x} \in X\}$ for this problem.

[6.3] Let $\phi(\mathbf{x}, \mathbf{y})$ be a continuous function defined for $\mathbf{x} \in X \subseteq R^n$ and $\mathbf{y} \in Y \subseteq R^m$. Show that

$$\sup_{\mathbf{y}\in Y} \inf_{\mathbf{x}\in X} \phi(\mathbf{x},\mathbf{y}) \le \inf_{\mathbf{x}\in X} \sup_{\mathbf{y}\in Y} \phi(\mathbf{x},\mathbf{y}).$$

[6.4] Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i = 1,\dots, m$, $h_i(\mathbf{x}) = 0$ for $i = 1,\dots, \ell$, and $\mathbf{x} \in X$, and let v: $R^{m+\ell} \to R$ be the perturbation function defined by (6.9). Assuming that f and $\mathbf{g}$ are convex, $\mathbf{h}$ is affine, and that X is a convex set, show that v is a convex function.

[6.5] For the problem of Example 6.3.5, sketch the perturbation function v defined by (6.9), and comment on the existence of a saddle point solution.

[6.6] Let f: $R^n \to R$ be a concave function, and let $\partial f(\bar{\mathbf{x}})$ be the subdifferential of f at any $\bar{\mathbf{x}} \in R^n$. Show that the directional derivative of f at $\bar{\mathbf{x}}$ in the direction $\mathbf{d}$ is given by $f'(\bar{\mathbf{x}}; \mathbf{d}) = \inf\{\xi^t \mathbf{d} : \xi \in \partial f(\bar{\mathbf{x}})\}$. What is the corresponding result if f is a convex function?

[6.7] Consider the discrete optimization Problem DP: Minimize $\{\mathbf{c}^t\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{D}\mathbf{x} = \mathbf{d}$, $\mathbf{x} \in X\}$, where X is some compact discrete set, and assume that the problem is feasible. Define $\theta(\pi) = \min\{\mathbf{c}^t\mathbf{x} + \pi^t(\mathbf{A}\mathbf{x}-\mathbf{b}) : \mathbf{D}\mathbf{x} = \mathbf{d},\ \mathbf{x} \in X\}$ for any $\pi \in R^m$, where $\mathbf{A}$ is $m \times n$. Show that $\max\{\theta(\pi) : \pi \in R^m\} = \min\{\mathbf{c}^t\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \in \operatorname{conv}\{\mathbf{x} \in X : \mathbf{D}\mathbf{x} = \mathbf{d}\}\}$, where conv $\{\cdot\}$ denotes the convex hull operation. Use this result to interpret the duality gap that might exist between DP and the Lagrangian dual problem stated.

[6.8] Consider Problem DP given in Exercise 6.7, and rewrite this problem as minimize $\{\mathbf{c}^t\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{D}\mathbf{y} = \mathbf{d},\ \mathbf{x} = \mathbf{y},\ \mathbf{x} \in X,\ \mathbf{y} \in Y\}$, where Y is a copy of X in which the $\mathbf{x}$-variables have been replaced by a set of matching $\mathbf{y}$-variables. Formulate a Lagrangian dual function $\bar{\theta}(\mu) = \min\{\mathbf{c}^t\mathbf{x} + \mu^t(\mathbf{x}-\mathbf{y}) : \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{D}\mathbf{y} = \mathbf{d},\ \mathbf{x} \in X,\ \mathbf{y} \in Y\}$. Show that $\max\{\bar{\theta}(\mu) : \mu \in R^n\} \ge \max\{\theta(\pi) : \pi \in R^m\}$, where θ is defined in Exercise 6.7. Discuss this result in relation to the respective partial convex hulls corresponding to θ and $\bar{\theta}$ as presented in Section 6.4 and Exercise 6.7.

[6.9] Consider the pair of primal and dual Problems P and D stated in Section 6.1, and assume that the Lagrangian dual function θ is differentiable. Given $(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \in R^{m+\ell}$, $\bar{\mathbf{u}} \ge 0$, let $\nabla\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})^t = [\mathbf{g}(\bar{\mathbf{x}})^t, \mathbf{h}(\bar{\mathbf{x}})^t]$, and define $\hat{g}_i(\bar{\mathbf{x}}) = g_i(\bar{\mathbf{x}})$ if $\bar{u}_i > 0$ and $\hat{g}_i(\bar{\mathbf{x}}) = \max\{0, g_i(\bar{\mathbf{x}})\}$ if $\bar{u}_i = 0$, for $i = 1,\dots, m$. If $(\mathbf{d}_u, \mathbf{d}_v) \equiv [\hat{\mathbf{g}}(\bar{\mathbf{x}}),$

$h(\bar{x})] \neq (0, 0)$, then show that $(\mathbf{d}_u, \mathbf{d}_v)$ is a feasible ascent direction of θ at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Hence, discuss how θ can be maximized in the direction $(\mathbf{d}_u, \mathbf{d}_v)$ via the one-dimensional problem to $\text{maximize}_\lambda \{\theta(\bar{\mathbf{u}} + \lambda \mathbf{d}_u, \bar{\mathbf{v}} + \lambda \mathbf{d}_v) : \bar{\mathbf{u}} + \lambda \mathbf{d}_u \geq \mathbf{0}, \lambda \geq 0\}$. On the other hand, if $(\mathbf{d}_u, \mathbf{d}_v) = (0, 0)$, then show that $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solves D.

Consider the problem to minimize $x_1^2 + x_2^2$ subject to $g_1(\mathbf{x}) = -x_1 - x_2 + 4 \leq 0$ and $g_2(\mathbf{x}) = x_1 + 2x_2 - 8 \leq 0$. Illustrate the *gradient method* presented above by starting at the dual solution $(u_1, u_2) = (0, 0)$ and verifying that after one iteration of this method, an optimal solution is obtained in this case.

[6.10] Consider the problem to minimize $x_1^2 + x_2^2$ subject to $x_1 + x_2 - 4 \geq 0$ and $x_1, x_2 \geq 0$.

a. Verify that the optimal solution is $\bar{\mathbf{x}} = (2,2)^t$ with $f(\bar{\mathbf{x}}) = 8$.
b. Letting $X = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\}$, write the Lagrangian dual problem. Show that the dual function is $\theta(u) = -u^2/2 - 4u$. Verify that there is no duality gap for this problem.
c. Solve the dual problem by the cutting plane algorithm of Section 6.4. Start with $\mathbf{x} = (3,3)^t$.
d. Show that θ is differentiable everywhere, and solve the problem using the gradient method of Exercise 6.9.

[6.11] Consider the following problem:

$$\begin{aligned} \text{Minimize } & (x_1 - 2)^2 + (x_2 - 6)^2 \\ \text{subject to } & x_1^2 - x_2 \leq 0 \\ & -x_1 \leq 1 \\ & 2x_1 + 3x_2 \leq 18 \\ & x_1, \quad x_2 \geq 0. \end{aligned}$$

a. Find the optimal solution geometrically, and verify it by using the KKT conditions.
b. Formulate the dual problem in which $X = \{(x_1, x_2) : 2x_1 + 3x_2 \leq 18, x_1, x_2 \geq 0\}$.
c. Perform three iterations of the cutting plane algorithm described in Section 6.4, starting with $(u_1, u_2) = (0, 0)$. Describe the perturbed optimization problems corresponding to the generated primal infeasible points. Also identify the primal feasible solutions generated by the algorithm.

[6.12] In reference to Exercise 6.11, perform three iterations of the gradient method of Exercise 6.9 and compare the results with those obtained by the cutting plane algorithm.

[6.13] Consider the following problem:

$$\begin{aligned}
\text{Maximize } & 3x_1 + 6x_2 + 2x_3 + 4x_4 \\
\text{subject to } & x_1 + x_2 + x_3 + x_4 \le 12 \\
& -x_1 + x_2 + 2x_4 \le 4 \\
& x_1 + x_2 \le 12 \\
& x_2 \le 4 \\
& x_3 + x_4 \le 6 \\
& x_1, x_2, x_3, x_4 \ge 0.
\end{aligned}$$

a. Formulate the dual problem in which $X = \{(x_1, x_2, x_3, x_4): x_1 + x_2 \le 12, x_2 \le 4, x_3 + x_4 \le 6; x_1, x_2, x_3, x_4 \ge 0\}$.

b. Starting from the point (0, 0), solve the Lagrangian dual problem by optimizing along the direction of steepest ascent as discussed in Exercise 6.9.

c. At optimality of the dual, find the optimal primal solution.

[6.14] Consider the primal Problem P discussed in Section 6.1. Introducing the slack vector **s**, the problem can be formulated as follows:

$$\begin{aligned}
\text{Minimize } & f(\mathbf{x}) \\
\text{subject to } & \mathbf{g}(\mathbf{x}) + \mathbf{s} = \mathbf{0} \\
& \mathbf{h}(\mathbf{x}) = \mathbf{0} \\
& (\mathbf{x}, \mathbf{s}) \in X',
\end{aligned}$$

where $X' = \{(\mathbf{x}, \mathbf{s}) : \mathbf{x} \in X, \mathbf{s} \ge \mathbf{0}\}$. Formulate the dual of the above problem and show that it is equivalent to the dual problem discussed in Section 6.1.

[6.15] Consider the following problem:

$$\begin{aligned}
\text{Maximize } & 3x_1 + 2x_2 + x_3 \\
\text{subject to } & 2x_1 + x_2 - x_3 \le 2 \\
& x_1 + 2x_2 \le 4 \\
& x_3 \le 3 \\
& x_1, x_2, x_3 \ge 0.
\end{aligned}$$

a. Find explicitly the dual function, where $X = \{(x_1, x_2, x_3): 2x_1 + x_2 - x_3 \le 2; x_1, x_2, x_3 \ge 0\}$.

b. Repeat Part a for $X = \{(x_1, x_2, x_3): x_1 + 2x_2 \le 4; x_1, x_2, x_3 \ge 0\}$.

c. In Parts a and b, note that the difficulty in evaluating the dual function at a given point depends on which constraints are handled via the set X. Propose some general guidelines that could be used in selecting the set X to make the solution easier.

[6.16] Consider the problem to minimize e^{-2x} subject to $-x \le 0$.

a. Solve the above primal problem.

b. Letting $X = R$ find the explicit form of the Lagrangian dual function, and solve the dual problem.

[6.17] Consider the problem to minimize x_1 subject to $x_1^2 + x_2^2 = 4$. Derive the dual function explicitly, and verify its concavity. Find the optimal solutions to both the primal and dual problems, and compare their objective values.

[6.18] Under the assumptions of Theorem 6.2.5, suppose that $\bar{\mathbf{x}}$ is an optimal solution to the primal problem and that f and $\mathbf{g}$ are differentiable at $\bar{\mathbf{x}}$. Show that there exists a vector $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ such that

$$\left[\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} \bar{v}_i \nabla h_i(\bar{\mathbf{x}})\right]^t (\mathbf{x} - \bar{\mathbf{x}}) \geq 0 \qquad \text{for each } \mathbf{x} \in X$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots,m$$

$$\bar{\mathbf{u}} \geq 0.$$

Show that these conditions reduce to the KKT conditions if X is open.

[6.19] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{x} \in X$. Theorem 6.2.4 shows that the primal and dual objective values are equal at optimality under the assumptions that f, $\mathbf{g}$, and X are convex and that the constraint qualification $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ for some $\hat{\mathbf{x}} \in X$ holds true. Suppose that the convexity assumptions on f and $\mathbf{g}$ are replaced by continuity of f and $\mathbf{g}$ and that X is assumed to be convex and compact. Does the result of the theorem hold true? Prove or give a counterexample.

[6.20] In the proof of Lemma 6.2.3, show that the set Λ is convex.

[6.21] Prove the following saddle point optimality condition. Let X be a nonempty convex set in R^n, and let f: $R^n \to R$, $\mathbf{g}$: $R^n \to R^m$ be convex and $\mathbf{h}$: $R^n \to R^\ell$ be affine. If $\bar{\mathbf{x}}$ is an optimal solution to the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{x} \in X$, then there exist $(\bar{u}_0, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \neq \mathbf{0}$, $(\bar{u}_0, \bar{\mathbf{u}}) \geq \mathbf{0}$ such that

$$\phi(\bar{u}_0, \mathbf{u}, \mathbf{v}, \bar{\mathbf{x}}) \leq \phi(\bar{u}_0, \bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\mathbf{x}}) \leq \phi(\bar{u}_0, \bar{\mathbf{u}}, \bar{\mathbf{v}}, \mathbf{x})$$

for all $\mathbf{u} \geq 0$, $\mathbf{v} \in R^\ell$ and $\mathbf{x} \in X$, where $\phi(u_0, \mathbf{u}, \mathbf{v}, \mathbf{x}) = u_0 f(\mathbf{x}) + \mathbf{u}^t \mathbf{g}(\mathbf{x}) + \mathbf{v}^t \mathbf{h}(\mathbf{x})$.

[6.22] Let P and D be the primal and dual nonlinear programs stated in Section 6.1, and denote $\mathbf{w} = (\mathbf{u}, \mathbf{v})$. Suppose that $\bar{\mathbf{w}}$ solves D. If there exists a saddle point solution to P and if $\bar{\mathbf{x}}$ solves uniquely for $\theta(\bar{\mathbf{w}})$, then show that $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ is such a saddle point solution. Correspondingly, if θ is differentiable at $\bar{\mathbf{w}}$, and if $\bar{\mathbf{x}}$ (uniquely) solves for θ at $\bar{\mathbf{w}}$, then show that $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ is a saddle point solution. (In particular, this shows that if Problem P has no saddle point solution, then θ cannot be differentiable at optimality.)

[6.23] Consider the following problem:

$$\begin{aligned}
\text{Minimize } & -2x_1 + 2x_2 + x_3 - 3x_4 \\
\text{subject to } & x_1 + x_2 + x_3 + x_4 \le 8 \\
& x_1 - 2x_3 + 4x_4 \le 2 \\
& x_1 + x_2 \le 8 \\
& x_3 + 2x_4 \le 6 \\
& x_1, \quad x_2, \quad x_3, \quad x_4 \ge 0.
\end{aligned}$$

Let $X = \{(x_1, x_2, x_3, x_4): x_1 + x_2 \le 8,\ x_3 + 2x_4 \le 6;\ x_1, x_2, x_3, x_4 \ge 0\}$.

a. Find the function θ explicitly.
b. Verify that θ is differentiable at (4, 0), and find $\nabla\theta(4, 0)$.
c. Verify that $\nabla\theta(4,0)$ is an infeasible direction, and find an improving feasible direction.
d. Starting from (4, 0), maximize θ in the direction obtained in Part c.

[6.24] Consider the following problem:

$$\begin{aligned}
\text{Minimize } & 2x_1 + x_2 \\
\text{subject to } & x_1 + 2x_2 \le 8 \\
& 2x_1 + 3x_2 \le 6 \\
& x_1, \quad x_2 \ge 0 \\
& x_1, \quad x_2 \quad \text{integers.}
\end{aligned}$$

Let $X = \{(x_1, x_2): 2x_1 + 3x_2 \le 6,\ x_1,\ x_2 \ge 0 \text{ and integer}\}$. At $u = 2$, is θ differentiable? If not, characterize its ascent directions.

[6.25] Construct a numerical problem in which a subgradient of the dual function is not an ascent direction. Is it possible that the collection of subgradients and the cone of ascent directions are disjoint at a nonoptimal solution? (*Hint*: Consider the shortest subgradient.)

[6.26] Suppose that $\theta: R^m \to R$ is concave.

a. Show that θ achieves its maximum at $\bar{\mathrm{u}}$ if and only if

$$\max\{\theta'(\bar{\mathrm{u}};\mathrm{d}) : \|\mathrm{d}\| \le 1\} = 0.$$

b. Show that θ achieves its maximum over the region $U = \{\mathrm{u} : \mathrm{u} \ge 0\}$ at $\bar{\mathrm{u}}$ if and only if

$$\max\{\theta'(\bar{\mathrm{u}};\mathrm{d}) : \mathrm{d} \in D, \|\mathrm{d}\| \le 1\} = 0,$$

where D is the cone of feasible directions of U at $\bar{\mathrm{u}}$. (Note that the above results can be used as stopping criteria for maximizing the Lagrangian dual function.)

[6.27] Consider the problem to minimize x subject to $g(x) \le 0$ and $x \in X = \{x : x \ge 0\}$. Derive the explicit form of the Lagrangian dual function, and determine the collection of subgradients at $u = 0$ for each of the following cases:

a. $g(x) = \begin{cases} -2/x & \text{for } x \ne 0 \\ 0 & \text{for } x = 0. \end{cases}$

b. $g(x) = \begin{cases} -2/x & \text{for } x \ne 0 \\ -1 & \text{for } x = 0. \end{cases}$

c. $g(x) = \begin{cases} 2/x & \text{for } x \ne 0 \\ 1 & \text{for } x = 0. \end{cases}$

[6.28] Consider the cutting plane method described in Section 6.4, and suppose that each time the master program objective value strictly increases, we delete all the constraints of the type $z \le f(\mathbf{x}_j) + \mathbf{u}^t \mathbf{g}(\mathbf{x}_j) + \mathbf{v}^t \mathbf{h}(\mathbf{x}_j)$ that are nonbinding at optimality. If X has a finite number of elements, show that this modified algorithm will converge finitely. Give some alternative conditions under which such a constraint deletion will assure convergence of the algorithm.

[6.29] Consider the following problem, in which X is a compact polyhedral set and f is a concave function:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \in X. \end{aligned}$$

a. Formulate the Lagrangian dual problem.
b. Show that the dual function is concave and piecewise linear.
c. Characterize the subgradients, the ascent directions, and the steepest ascent direction for the dual function.
d. Generalize the result in Part b to the case where X is not compact.

[6.30] Consider the pair of primal and dual Problems P and D stated in Section 6.1, and suppose that the Lagrangian dual function θ is not necessarily differentiable. Given $\bar{\mathbf{w}} = (\bar{\mathbf{u}}, \bar{\mathbf{v}}) \in R^{m+\ell}$, $\bar{\mathbf{u}} \ge \mathbf{0}$, let $\xi_1, \ldots, \xi_p$, $p \ge 1$, be some known collection of subgradients of θ at $\bar{\mathbf{w}}$. Consider the problem to maximize $\{\varepsilon : \mathbf{d}^t \xi_j \ge \varepsilon \text{ for } j = 1, \ldots, p,\ -1 \le d_i \le 1 \text{ for } i = 1, \ldots, m+\ell, \text{ with } d_i \ge 0 \text{ if } \bar{u}_i = 0\}$. Let $(\varepsilon, \bar{\mathbf{d}})$ solve this problem. If $\varepsilon = 0$, show that $\bar{\mathbf{w}}$ solves D. Otherwise, solve the problem to maximize $\{\bar{\mathbf{d}}^t \xi : \xi \in \partial\theta(\bar{\mathbf{w}})\}$, and let ξ_{p+1} be an optimum. If $\bar{\mathbf{d}}^t \xi_{p+1} > 0$, then show that $\bar{\mathbf{d}}$ is an ascent direction along which θ can be maximized by solving $\max\{\theta(\bar{\mathbf{w}} + \lambda \bar{\mathbf{d}}) : \bar{u}_i + \lambda \bar{d}_i \ge 0 \text{ for } i = 1, \ldots, m,\ \lambda \ge 0\}$, and the process can then be repeated. Otherwise, if $\bar{\mathbf{d}}^t \xi_{p+1} \le 0$, then increment p by 1 and re-solve the direction-finding problem given above. Discuss the possible computational difficulties associated with this scheme. How would you implement the various steps if all functions were affine and X was a nonempty polytope? Illustrate this using the example to minimize $x_1 - 4x_2$ subject to $-x_1 - x_2 + 2 \le 0$, $x_2 - 1 \le 0$ and $\mathbf{x} \in X = \{\mathbf{x} : 0 \le x_1 \le 3, 0 \le x_2 \le 3\}$, starting at the point $(u_1, u_2) = (0, 4)$.

[6.31] Consider the linear program to minimize $\mathbf{c}^t\mathbf{x}$ subject to $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$. Write the dual problem. Show that the dual of the dual problem is equivalent to the primal problem.

[6.32] Consider the following problem:

$$\begin{array}{llcrcrcl}
\text{Minimize} & -2x_1 & - & 2x_2 & - & x_3 & & \\
\text{subject to} & 2x_1 & + & x_2 & + & x_3 & \le & 8 \\
& 3x_1 & - & 2x_2 & + & 3x_3 & \le & 3 \\
& x_1 & + & x_2 & & & \le & 5 \\
& x_1, & & x_2, & & x_3 & \ge & 0.
\end{array}$$

Solve the primal problem by the simplex method. At each iteration identify the dual variables from the simplex tableau. Show that the dual variables satisfy the complementary slackness conditions but violate the dual constraints. Verify that dual feasibility is attained at termination.

[6.33] Consider the primal and dual linear programming problems discussed in Section 6.6. Show directly using Farkas's lemma that if the primal is inconsistent and the dual admits a feasible solution, the dual has an unbounded objective value.

[6.34] In Section 6.3 we showed that the shortest subgradient ξ of θ at $\bar{\mathbf{u}}$ is the steepest ascent direction. The following modification of ξ is proposed to maintain feasibility:

$$\bar{\xi}_i = \begin{cases} \max\{0, \xi_i\} & \text{if } \bar{u}_i = 0 \\ \xi_i & \text{if } \bar{u}_i \ge 0. \end{cases}$$

Is $\bar{\xi}$ an ascent direction? Is it the direction of steepest ascent with the added nonnegativity restriction? Prove or give a counterexample.

[6.35] Suppose that the shortest subgradient $\bar{\xi}$ of θ at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ is not equal to zero. Show that there exists an $\varepsilon > 0$ such that $\left\|\xi - \bar{\xi}\right\| < \varepsilon$ implies that ξ is an ascent direction of θ at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. (From this exercise, if an iterative procedure is used to find $\bar{\xi}$, it would find an ascent direction after a sufficient number of iterations.)

[6.36] Consider a singly constrained problem to minimize $f(\mathbf{x})$ subject to $g(\mathbf{x}) \le 0$ and $\mathbf{x} \in X$, where X is a compact set. The Lagrangian dual problem is to maximize $\theta(u)$ subject to $u \ge 0$, where $\theta(u) = \inf\{f(\mathbf{x}) + ug(\mathbf{x}) : \mathbf{x} \in X\}$.

a. Let $\hat{u} \ge 0$, and let $\hat{\mathbf{x}} \in X(\hat{u})$. Show that if $g(\hat{\mathbf{x}}) > 0$, then $\bar{u} > \hat{u}$, and if $g(\hat{\mathbf{x}}) < 0$, then $\bar{u} < \hat{u}$, where $\bar{u}$ is an optimal solution to the Lagrangian dual.
b. Use the result of Part a to find an interval $[a, b]$ that contains all the optimal solutions to the dual problem or else to conclude that the dual problem is unbounded.
c. Now consider the problem to maximize $\theta(u)$ subject to $a \le u \le b$. Suppose that the following scheme is used to solve the problem: Let $\bar{u} = (a + b)/2$, and let $\bar{\mathbf{x}} \in X(\bar{u})$. If $g(\bar{\mathbf{x}}) > 0$, replace a by $\bar{u}$ and repeat the process. If $g(\bar{\mathbf{x}}) < 0$, replace b by $\bar{u}$ and repeat the process. If $g(\bar{\mathbf{x}}) = 0$, stop; $\bar{u}$ is an optimal dual solution. Show that the procedure converges to an optimal solution, and illustrate by solving the dual of the following problem:

$$\begin{aligned} \text{Minimize} \quad & 2x_1^2 + x_2^2 \\ \text{subject to} \quad & -x_1 - 2x_2 + 2 \le 0. \end{aligned}$$

d. An alternative approach to solving the problem to maximize $\theta(u)$ subject to $a \le u \le b$ is to specialize the tangential approximation method discussed in Section 6.4. Show that at each iteration only two supporting hyperplanes need be considered, and that the method could be stated as follows: Let $\mathbf{x}_a \in X(a)$ and $\mathbf{x}_b \in X(b)$. Let $\bar{u} = [f(\mathbf{x}_a) - f(\mathbf{x}_b)]/[g(\mathbf{x}_b) - g(\mathbf{x}_a)]$. If $\bar{u} = a$ or $\bar{u} = b$, stop; $\bar{u}$ is an optimal solution to the dual problem. Otherwise, let $\bar{\mathbf{x}} \in X(\bar{u})$. If $g(\bar{\mathbf{x}}) > 0$, replace a by $\bar{u}$ and repeat the process. If $g(\bar{\mathbf{x}}) < 0$, replace b by $\bar{u}$ and repeat the process. If $g(\bar{\mathbf{x}}) = 0$, stop; $\bar{u}$ is an optimal dual solution. Show that the procedure converges to an optimal solution, and illustrate by solving the problem in Part c.

[6.37] Consider the primal and Lagrangian dual problems discussed in Section 6.1. Let $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ be an optimal solution to the dual problem. Given $(\mathbf{u}, \mathbf{v})$, suppose that $\bar{\mathbf{x}} \in X(\mathbf{u}, \mathbf{v})$, as defined in Section 6.3. Show that there exists a $\delta > 0$ such that $\|(\bar{\mathbf{u}}, \bar{\mathbf{v}}) - (\mathbf{u}, \mathbf{v}) - \lambda[\mathbf{g}(\bar{\mathbf{x}}), \mathbf{h}(\bar{\mathbf{x}})]\|$ is a nonincreasing function of λ over the interval $[0, \delta]$. Interpret the result geometrically, and illustrate by the following problem, in which $(u_1, u_2) = (3, 1)$ are the dual variables corresponding to the first two constraints:

$$\begin{aligned} \text{Minimize } & -2x_1 - 2x_2 - 5x_3 \\ \text{subject to } & x_1 + x_2 + x_3 \le 10 \\ & x_1 + 2x_3 \ge 6 \\ & x_1, \quad x_2, \quad x_3 \le 3 \\ & x_1, \quad x_2, \quad x_3 \ge 0. \end{aligned}$$

[6.38] From Exercise 6.37 it is clear that moving a small step in the direction of any subgradient leads us closer to an optimal dual solution. Consider the following algorithm for maximizing the dual of the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{x} \in X$.

Main Step

Given $\mathbf{v}_k$, let $\mathbf{x}_k \in X(\mathbf{v}_k)$. Let $\mathbf{v}_{k+1} = \mathbf{v}_k + \lambda \mathbf{h}(\mathbf{x}_k)$, where $\lambda > 0$ is a small scalar. Replace k by $k + 1$ and repeat the main step.

a. Discuss some possible ways of choosing a suitable step size λ. Do you see any advantages in reducing the step size during later iterations? If so, propose a scheme for doing that.
b. Does the dual function necessarily increase from one iteration to another? Discuss.
c. Devise a suitable termination criterion.
d. Apply the above algorithm, starting from $\mathbf{v} = (1, 2)^t$, to solve the following problem:

$$\begin{aligned} \text{Minimize } & x_1^2 + x_2^2 + 2x_3 \\ \text{subject to } & x_1 + x_2 + x_3 = 6 \\ & -x_1 + x_2 + x_3 = 4. \end{aligned}$$

(This procedure, with a suitable step size selection rule, is referred to as a *subgradient optimization* technique. See Chapter 8 for further details.)

[6.39] Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$, $\mathbf{x} \in X$.

a. In Exercise 6.38, a subgradient optimization technique was discussed for the equality case. Modify the procedure for the above inequality-constrained problem. [*Hint*: Given $\mathbf{u}$, let $\mathbf{x} \in X(\mathbf{u})$. Replace $g_i(\mathbf{x})$ by $\max\{0, g_i(\mathbf{x})\}$ for each i with $u_i = 0$.]
b. Illustrate the procedure given in Part a by solving the problem in Exercise 6.13 starting from $\mathbf{u} = (0, 0)^t$.
c. Extend the subgradient optimization technique to handle both equality and inequality constraints.

[6.40] Consider the problems to find

$$\min_{\mathbf{x}\in X}\max_{\mathbf{y}\in Y}\phi(\mathbf{x},\mathbf{y}) \quad \text{and} \quad \max_{\mathbf{y}\in Y}\min_{\mathbf{x}\in X}\phi(\mathbf{x},\mathbf{y}),$$

where X and Y are nonempty compact convex sets in R^n and R^m, respectively, and ϕ is convex in $\mathbf{x}$ for any given $\mathbf{y}$ and concave in $\mathbf{y}$ for any given $\mathbf{x}$.

a. Show that $\min_{\mathbf{x}\in X}\max_{\mathbf{y}\in Y}\phi(\mathbf{x},\mathbf{y}) \geq \max_{\mathbf{y}\in Y}\min_{\mathbf{x}\in X}\phi(\mathbf{x},\mathbf{y})$ without any convexity assumptions.

b. Show that $\max_{\mathbf{y}\in Y}\phi(\cdot,\mathbf{y})$ is a convex function in $\mathbf{x}$ and that $\min_{\mathbf{x}\in X}\phi(\mathbf{x},\cdot)$ is a concave function in $\mathbf{y}$.

c. Show that $\min_{\mathbf{x}\in X}\max_{\mathbf{y}\in Y}\phi(\mathbf{x},\mathbf{y}) = \max_{\mathbf{y}\in Y}\min_{\mathbf{x}\in X}\phi(\mathbf{x},\mathbf{y})$.

(*Hint*: Use Part b and the necessary optimality conditions of Section 3.4.)

[6.41] Consider the following problem, in which X is a compact polyhedral set:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^t\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \in X. \end{aligned}$$

For a given dual vector $\mathbf{v}$, suppose that $\mathbf{x}_1,\ldots,\mathbf{x}_k$ are the extreme points in X that belong to $X(\mathbf{v})$ as defined in Section 6.3. Show that the extreme points of $\partial\theta(\mathbf{v})$ are contained in the set $\Lambda = \{\mathbf{A}\mathbf{x}_j - \mathbf{b} : j = 1,\ldots,k\}$. Give an example where the extreme points of $\partial\theta(\mathbf{v})$ form a proper subset of Λ.

[6.42] A company wants to plan its production rate of a certain item over the planning period $[0, T]$ such that the sum of its production and inventory costs is minimized. In addition, the known demand must be met, the production rate must fall in the acceptable interval $[\ell, u]$, the inventory must not exceed d, and it must be at least equal to b at the end of the planning period. The problem can be formulated as follows:

$$\begin{aligned} \text{Minimize} \quad & \int_0^T [c_1 x(t) + c_2 y^2(t)]\, dt \\ \text{subject to} \quad & x(t) = x_0 + \int_0^t [y(\tau) - z(\tau)]\, d\tau && \text{for } t \in [0,T] \\ & x(T) \geq b \\ & 0 \leq x(t) \leq d && \text{for } t \in (0,T) \\ & \ell \leq y(t) \leq u && \text{for } t \in (0,T), \end{aligned}$$

where

$x(t)$ = inventory at time t

$y(t)$ = production rate at time t

$z(t)$ = known demand rate at time t

x_0 = known initial inventory

c_1, c_2 = known coefficients

a. Make the above control problem discrete as was done in Section 1.2, and formulate a suitable Lagrangian dual problem.
b. Make use of the results of this chapter to develop a scheme for solving the primal and dual problems.
c. Apply your algorithm to the following data: $T = 6$, $x_0 = 0$, $b = 4$, $c_1 = 1$, $c_2 = 2$, $\ell = 2$, $u = 5$, $d = 6$, and $z(t) = 4$ over $[0, 4]$ and $z(t) = 3$ over $(4, 6]$.

[6.43] Consider the following warehouse location problem. We are given destinations 1,..., k, where the known demand for a certain product at destination j is d_j. We are also given m possible sites for building warehouses. If we decide to build a warehouse at site i, its capacity has to be s_i, and it incurs a fixed cost f_i. The unit shipping cost from warehouse i to destination j is c_{ij}. The problem is to determine how many warehouses to build, where to locate them, and what shipping patterns to use so that the demand is satisfied and the total cost is minimized. The problem can be stated mathematically as follows:

$$\begin{array}{lll} \text{Minimize} & \sum_{i=1}^{m}\sum_{j=1}^{k} c_{ij}x_{ij} + \sum_{i=1}^{m} f_i y_i & \\ \text{subject to} & \sum_{j=1}^{k} x_{ij} \le s_i y_i & \text{for } i = 1,\dots,m \\ & \sum_{i=1}^{m} x_{ij} \ge d_j & \text{for } j = 1,\dots,k \\ & 0 \le x_{ij} \le y_i \min\{s_i, d_j\} & \text{for } i = 1,\dots,m;\ j = 1,\dots,k \\ & y_i = 0 \text{ or } 1 & \text{for } i = 1,\dots,m. \end{array}$$

a. Formulate a suitable Lagrangian dual problem. Explain the utility of the upper bound imposed on x_{ij}.
b. Make use of the results of this chapter to devise a special scheme for maximizing the dual of the warehouse location problem.
c. Illustrate by a small numerical example.

[6.44] Consider the (primal) quadratic program PQP: Minimize $\{\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{D}\mathbf{x} : \mathbf{A}\mathbf{x} \ge \mathbf{b}\}$, where $\mathbf{D}$ is an $n \times n$ symmetric matrix and $\mathbf{A}$ is $m \times n$. Let W be an arbitrary set such that $\{\mathbf{w} : \mathbf{A}\mathbf{w} \ge \mathbf{b}\} \subseteq W$, and consider Problem EDQP: Minimize $\{\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{w}^t\mathbf{D}\mathbf{w} : \mathbf{A}\mathbf{x} \ge \mathbf{b},\ \mathbf{D}\mathbf{w} = \mathbf{D}\mathbf{x},\ \mathbf{w} \in W\}$.

a. Show that PQP and EPQP are equivalent in the sense that if $\mathbf{x}$ is feasible to PQP, $(\mathbf{x}, \mathbf{w})$ with $\mathbf{w} = \mathbf{x}$ is feasible to EDQP with the same objective value; and conversely, if $(\mathbf{x}, \mathbf{w})$ is feasible to EPQP, $\mathbf{x}$ is feasible to PQP with the same objective value.

b. Construct a Lagrangian dual LD: Maximize $\{\theta(\mathbf{y})\}$, where $\theta(\mathbf{y}) = \min\{(\mathbf{c}+\mathbf{D}\mathbf{y})^t\mathbf{x} + (1/2)\mathbf{w}^t\mathbf{D}\mathbf{w} - \mathbf{y}^t\mathbf{D}\mathbf{w} : \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{w} \in W\}$. Show that equivalently, we have

$$\text{LD:} \quad \sup\{\mathbf{b}^t\mathbf{u} - (1/2)\mathbf{y}^t\mathbf{D}\mathbf{y} + \phi(\mathbf{y}) : \mathbf{A}^t\mathbf{u} - \mathbf{D}\mathbf{y} = \mathbf{c}, \mathbf{u} \geq \mathbf{0}\},$$

where $\phi(\mathbf{y}) = \inf\{(1/2)(\mathbf{y}-\mathbf{w})^t\mathbf{D}(\mathbf{y}-\mathbf{w}) : \mathbf{w} \in W\}$.

c. Show that if $\mathbf{D}$ is positive semidefinite and $W \equiv R^n$, $\phi(\mathbf{y}) = 0$ for all $\mathbf{y}$, and LD reduces to Dorn's dual program given in (6.29). On the other hand, if $\mathbf{D}$ is not positive semidefinite and $W \equiv R^n$, $\phi(\mathbf{y}) = -\infty$ for all $\mathbf{y}$. Furthermore, if PQP has an optimum of objective value v_p, and if $W = \{\mathbf{w} : \mathbf{A}\mathbf{w} \geq \mathbf{b}\}$, show that the optimum value of LD is also v_p. What does this suggest regarding the formulation of LD for nonconvex situations?

d. Illustrate Part c using the problem to minimize $\{x_1x_2 : x_1 \geq 0$ and $x_2 \geq 0\}$. (This exercise is based on Sherali [1993].)

[6.45] Consider the dual quadratic program given by (6.30). Describe a gradient-based maximization scheme for this problem, following Exercise 6.9. Can you anticipate any computational difficulties? (*Hint*: See Chapter 8.) Illustrate by using the following quadratic programming problem:

$$\begin{aligned}
\text{Minimize } & 3x_1^2 + 2x_2^2 - 2x_1x_2 - 3x_1 - 4x_2 \\
\text{subject to } & 2x_1 + 3x_2 \leq 6 \\
& -x_1 + 2x_2 \leq 2 \\
& x_1, \quad x_2 \geq 0.
\end{aligned}$$

At each iteration, identify the corresponding primal infeasible point as well as the primal feasible point. Develop a suitable measure of infeasibility and check its progress. Can you draw any general conclusions?

[6.46] Let X and Y be nonempty sets in R^n, and let $f, g: R^n \to R$. Consider the *conjugate functions* f^* and g^* defined as follows:

$$f^*(\mathbf{u}) = \inf\{f(\mathbf{x}) - \mathbf{u}^t\mathbf{x} : \mathbf{x} \in X\}$$

$$g^*(\mathbf{u}) = \sup\{g(\mathbf{x}) - \mathbf{u}^t\mathbf{x} : \mathbf{x} \in Y\}.$$

a. Interpret f^* and g^* geometrically.

b. Show that f^* is concave over X^* and g^* is convex over Y^*, where $X^* = \{u : f^*(u) > -\infty\}$ and $Y^* = \{u : g^*(u) < \infty\}$.

c. Prove the following *conjugate weak duality theorem*:

$$\inf\{f(\mathbf{x}) - g(\mathbf{x}) : \mathbf{x} \in X \cap Y\} \geq \sup\{f^*(\mathbf{u}) - g^*(\mathbf{u}) : \mathbf{u} \in X^* \cap Y^*\}.$$

d. Now suppose that f is convex, g is concave, $\text{int } X \cap \text{int } Y \neq \emptyset$, and $\inf\{f(\mathbf{x}) - g(\mathbf{x}) : \mathbf{x} \in X \cap Y\}$ is finite. Show that equality in Part c above holds true and that $\sup\{f^*(\mathbf{u}) - g^*(\mathbf{u}) : \mathbf{u} \in X^* \cap Y^*\}$ is achieved.

e. By suitable choices of f, g, X, and Y, formulate a nonlinear programming problem as follows:

$$\begin{aligned} &\text{Minimize} \quad f(\mathbf{x}) - g(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{x} \in X \cap Y. \end{aligned}$$

What is the form of the conjugate dual problem? Devise some strategies for solving the dual problem.

Notes and References

The powerful results of duality in linear programming and the saddle point optimality criteria for convex programming sparked a great deal of interest in duality in nonlinear programming. Early results in this area include the work of Cottle [1963b], Dorn [1960a], Hanson [1961], Mangasarian [1962], Stoer [1963], and Wolfe [1961].

More recently, several duality formulations that enjoy many of the properties of linear dual programs have evolved. These include the Lagrangian dual problem, the conjugate dual problem, the surrogate dual problem, and the mixed Lagrangian and surrogate, or *composite dual*, problem. In this chapter we concentrated on the Lagrangian dual formulation because, in our judgment, it is the most promising formulation from a computational standpoint and also because the results of this chapter give the general flavor of the results that one would obtain using other duality formulations. Those interested in studying the subject of conjugate duality may refer to Fenchel [1949], Rockafellar [1964, 1966, 1968, 1969, 1970], Scott and Jefferson [1984, 1989], and Whinston [1967]. For the subject of surrogate duality, where the constraints are grouped into a single constraint by the use of Lagrangian multipliers, refer to Greenberg and Pierskalla [1970b]. Several authors have developed duality formulations that retain the symmetry between the primal and dual problems. The works of Cottle [1963b], Dantzig et al. [1965], Mangasarian and Ponstein [1965], and Stoer [1963] are in this class. For composite duality, see Karwan and Rardin [1979, 1980].

The reader will find the work of Geoffrion [1971b] and Karamardian [1967] excellent references on various duality formulations and their interrelationships. See Everett [1963], Falk [1967, 1969], and Lasdon [1968] for a further study on duality. The relationship between the Lagrangian duality formulation and other duality formulations is examined in Bazaraa et al. [1971b], Magnanti [1974], and Whinston [1967]. The economic interpretation of duality is covered by Balinski and Baumol [1968], Beckmann and Kapur [1972], Peterson [1970], and Williams [1970].

In Sections 6.1 and 6.2 the dual problem is presented and some of its properties are developed. As a by-product of the main duality theorem, we develop the saddle point optimality criteria for convex programs. These criteria were first developed by Kuhn and Tucker [1951]. For the related concept of min–max duality, see Mangasarian and Ponstein [1965], Ponstein [1965], Rockafellar [1968], and Stoer [1963]. For further discussions and illustrations of perturbation functions, see Geoffrion [1971b] and Minoux [1986]. Larsson and Patriksson [2003] provide a generalized set of near-saddle point optimality conditions and lay the foundation for Lagrangian-based heuristics. For some fundamental discussions and applications of Lagrangian relaxation/dual-based approaches for discrete problems, see Fisher [1981, 1985], Geoffrion [1974], and Shapiro [1979b]. Guignard and Kim [1987] discuss a useful concept of *Lagrangian decomposition* for exploiting special structures and formulating suitable Lagrangian duals for discrete and nonconvex problems, and Guignard [1998] discusses the value of adding additional constraints (cuts) in a Lagrangian relaxation framework that would have the potential for tightening the relaxation-based bound.

In Section 6.3 we examine several properties of the dual function. We characterize the collection of subgradients at any given point, and use that to determine both ascent directions and the steepest ascent direction. We show that the steepest ascent direction is the shortest subgradient. This result is essentially given by Demyanov [1968]. In Section 6.4 we use these properties to suggest several gradient-based or outer-linearization methods for maximizing the dual function. An accelerated version of the cutting plane method that ensures the generation of ascent directions is discussed by Hearn and Lawphongpanich [1989, 1990]. For a further study of this subject, see Bazaraa and Goode [1979], Demyanov [1968, 1971], Fisher et al. [1975], and Lasdon [1970]. For constraint deletion concepts in outer-linearization methods, see Eaves and Zangwill [1971] and Lasdon [1970]. There are other procedures for solving the dual problem. The cutting plane method discussed in Section 6.4 is a row generation procedure. In its dual form, it is precisely the column generation generalized programming method of Wolfe (see Dantzig [1963]). Another procedure is the subgradient optimization method, which is introduced briefly in Exercises 6.37, 6.38, and 6.39 and is discussed in more detail in Chapter 8. See Held et al. [1974] and Polyak [1967] for validation of subgradient optimization. For related work, see Bazaraa and Goode [1977, 1979], Bazaraa and Sherali [1981], Fisher et al. [1975], Held and Karp [1970], and Sherali et al. [2000].

One of the pioneering works for using the Lagrangian formulation to develop computational schemes is credited to Everett [1963]. Under certain conditions he showed how the primal solution could be retrieved. The result and its extensions are given in Section 6.5. For duality in quadratic programming, see Cottle [1963b], Dorn [1960a,b, 1961a], and Sherali [1993].

Chapter 11 Linear Complementary Problem, and Quadratic, Separable, Fractional, and Geometric Programming

In this chapter we introduce the linear complementary problem and develop some special procedures for solving quadratic, separable, and fractional programming problems. In each case, some variation of the simplex method is used as a solution procedure. For quadratic programming, the KKT system is solved by a complementary pivoting technique that could be used for the more general class of linear complementary problems. We also discuss a global optimization approach for this problem that can additionally handle quadratic constraints. (This approach can actually be extended to determine global optimal solutions to general *polynomial programming problems*—see Exercise 11.27.) For problems that are separable in the variables, we develop an approximation approach via a piecewise linearization, and the simplex method is used with a suitable restriction on basis entry to solve these problems. We also describe two simplex-based methods for solving linear fractional programs. Finally, we discuss geometric programming problems from a Lagrangian duality viewpoint. Such problems find varied applications in engineering design contexts.

Following is an outline of the chapter.

Section 11.1: Linear Complementary Problem We discuss mainly Lemke's algorithm for solving a linear complementary problem (LCP) and show its convergence in a finite number of iterations. Under suitable assumptions, the algorithm either stops with a complementary basic solution or concludes that the original system is inconsistent. Some comments for solving general LCPs are also provided.

Section 11.2: Convex and Nonconvex Quadratic Programming: Global Optimization Approaches We show that the KKT conditions for quadratic programs reduce to a linear complementary problem. The complementary pivoting algorithm is then used to solve the KKT system. Other approaches are discussed briefly, including a global optimization method based on the *reformulation–linearization/convexification technique* (RLT), with details being relegated to the exercises and the Notes and References section.

Section 11.3: Separable Programming Given a nonlinear programming problem whose objective and constraints are separable in the variables, each function can be approximated by a piecewise linear function using grid points. This is done in such a way that a slight modification of the simplex method can be used to solve the resulting problem. Under suitable convexity assumptions, the optimal objective value to the approximating problem can be made arbitrarily close to that of the original problem. Furthermore, we describe a scheme for generating grid points as needed.

Section 11.4: Linear Fractional Programming Linear fractional programming refers to problems of optimizing the ratio of two linear functions in the presence of linear constraints. We present two procedures for solving the problem. The first method is a simplified version of the convex-simplex method. The second method obtains an optimal solution by solving an equivalent linear program having an additional constraint and an additional variable.

Section 11.5: Geometric Programming This class of problems often arises in engineering applications. We present a technique for solving constrained polynomial geometric programming problems based on the use of Lagrangian duality concepts along with suitable transformations.

11.1 Linear Complementary Problem

In this section we briefly introduce the linear complementary problem and present a complementary pivoting algorithm for solving it. Problems of this type arise frequently in engineering applications, game theory, and economics. Also, as will be seen in Section 11.2, the KKT conditions for linear and quadratic programming problems can be written as a linear complementary problem, and hence, the algorithm presented in this section can be used to solve both linear and quadratic programming problems. Furthermore, the algorithm can be used to solve matrix game theory problems.

11.1.1 Definition

Let $\mathbf{M}$ be a given $p \times p$ matrix and let $\mathbf{q}$ be a given p-vector. The *linear complementary problem* (LCP) is to find vectors $\mathbf{w}$ and $\mathbf{z}$ such that

$$\mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q} \tag{11.1}$$

$$w_j \geq 0,\ z_j \geq 0 \qquad \text{for } j = 1,\ldots, p \tag{11.2}$$

$$w_j z_j = 0 \qquad \text{for } j = 1,\ldots, p \tag{11.3}$$

or to conclude that no such solution exists. Here (w_j, z_j) is a pair of *complementary variables*. A solution $(\mathbf{w}, \mathbf{z})$ to the above system is called a *complementary* *feasible solution.* Moreover, such a solution is a *complementary basic feasible solution* if (w, z) is a basic feasible solution to (11.1) and (11.2) with one variable of the pair (w_j, z_j) basic for each $j = 1,..., p$. Also, the restrictions (11.3) are sometimes referred to as *complementarity constraints.*

Let $\mathbf{e}_j$ denote a unit vector with a 1 in the jth position, and let $\mathbf{m}_j$ denote the jth column of **M** for $j = 1,..., p$. A cone spanned by any p vectors obtained by selecting one vector from each pair $\mathbf{e}_j$, and $-\mathbf{m}_j$ for $j = 1,..., p$, is called a *complementary cone* associated with the matrix **M** that defines the system (11.1)–(11.3). Note that there are 2^p such complementary cones, and that the above system has a solution if and only if **q** belongs to at least one such cone. Also, observe that if **q** belongs to a particular complementary cone and its generators constitute a basis, that is, they are linearly independent, then the corresponding solution is a complementary basic feasible solution, and vice versa. Furthermore, a square matrix **M** is called a *Q-matrix* if the corresponding system (11.1)–(11.3) has a solution for each $\mathbf{q} \in R^p$.

Using the concept of complementary cones to characterize a solution to linear complementary problems, we can cast (11.1)–(11.3) as an optimization problem in the following manner. Define a binary variable y_j to take on a value of zero or one accordingly as the variable w_j or z_j is permitted to be positive from the complementary pair (w_j, z_j) for each $j = 1,..., p$, and consider the following *mixed-integer zero-one bilinear programming problem* (BLP):

$$\text{BLP: Minimize } \left\{ \sum_{j=1}^{p} y_j w_j + (1-y_j) z_j : \right. \\ \left. \mathbf{w} - \mathbf{Mz} = \mathbf{q}, \mathbf{w} \geq \mathbf{0}, \mathbf{z} \geq \mathbf{0}, \quad \text{and } \mathbf{y} \text{ binary} \right\}. \tag{11.4}$$

Note that the objective function value for BLP is zero for any feasible solution if and only $y_j w_j = (1 - y_j) z_j = 0$ for each $j = 1,..., p$, since all the objective terms are nonnegative. Moreover, this happens at optimality if and only if $w_j z_j = 0$ for each $j = 1,..., p$ because of the binariness of y. Hence, a solution (w, z) is a solution to LCP if and only if it is part of an optimal solution to BLP having a zero objective function value.

Also, observe that we can relax the binary restrictions on the y variables in (11.4) equivalently to $\mathbf{0} \leq \mathbf{y} \leq \mathbf{e}$, where **e** is a vector of p ones. This follows since for any partial optimal solution $(\overline{\mathbf{w}}, \overline{\mathbf{z}})$ to BLP, the resulting problem to minimize $\{\sum_{j=1}^{p} [y_j \overline{w}_j + (1-y_j)\overline{z}_j] : \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}$ automatically yields a binary optimal solution for y. Hence, we can also consider (11.4) to be a (continuous) *bilinear programming problem* (also see Exercises 11.4 and 11.27), which is linear in y when (w, z) is fixed in value, and is linear in (w, z) when y is fixed in value. Because of the latter property, it follows that if LCP has a solution, then there exists a solution that is an extreme point of (11.1)–(11.2).

Moreover, using the foregoing characterization of LCP, it can also be cast as one of minimizing a concave objective function $h(\mathbf{y})$ subject to $\mathbf{0} \le \mathbf{y} \le \mathbf{e}$, where

$$h(\mathbf{y}) \equiv \min\{ \sum_{j=1}^{n} [y_j w_j + (1-y_j) z_j] : \mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q}, \mathbf{w} \ge \mathbf{0}, \text{ and } \mathbf{z} \ge \mathbf{0}\}$$

(see Exercise 11.9). Furthermore, assuming that the set defined by (11.1)–(11.2) is bounded, we can linearize BLP into a linear mixed-integer zero-one programming problem (see Exercises 11.4 and 11.6). Hence, we can solve general LCPs using available methods for bilinear programming, concave minimization, or linear integer programming problems. We refer the reader to the Notes and References section for such approaches, including certain specialized techniques, such as sequential linear programming methods or interior point approaches when **M** possesses particular structural properties. We now proceed to describe a popular simplex type of pivoting method for solving LCPs, which is guaranteed to work under certain nondegeneracy assumptions and when **M** satisfies certain properties. However, in practice, it is known to perform well even when these assumptions are violated.

Solving the Linear Complementary Problem

If **q** is nonnegative, then we immediately have a solution satisfying (11.1)–(11.3), by letting $\mathbf{w} = \mathbf{q}$ and $\mathbf{z} = \mathbf{0}$. If $\mathbf{q} \not\ge \mathbf{0}$, however, a new column **e** and an artificial variable are introduced, leading to the following system, where **e** is a vector of p ones.

$$\mathbf{w} - \mathbf{M}\mathbf{z} - \mathbf{e}z_0 = \mathbf{q} \tag{11.5}$$

$$z_0 \ge 0,\ w_j \ge 0,\ z_j \ge 0 \qquad \text{for } j = 1,\dots,p \tag{11.6}$$

$$w_j z_j = 0 \qquad \text{for } j = 1,\dots,p. \tag{11.7}$$

Letting z_0 = maximum $\{-q_i : 1 \le i \le p\}$, $\mathbf{z} = \mathbf{0}$, and $\mathbf{w} = \mathbf{q} + \mathbf{e}z_0$, we obtain a starting solution to the above system. Through a sequence of pivots, to be specified later, we attempt to drive the artificial variable z_0 to zero while satisfying (11.5)–(11.7), thus obtaining a solution to the linear complementary problem.

Consider the following definition of an almost complementary basic feasible solution and the definition of an adjacent almost complementary feasible solution. These definitions will be useful in both describing the algorithm and establishing its finite convergence.

11.1.2 Definition

Consider the system defined by (11.5)–(11.7). A feasible solution (w, z, z_0) to this system is called an *almost complementary basic feasible solution* if:

1. (w, z, z_0) is a basic feasible solution to (11.5) and (11.6).
2. Both w_s and z_s are nonbasic, for some $s \in \{1,..., p\}$.
3. z_0 is basic, and exactly one variable from each complementary pair (w_j, z_j) is basic, for $j = 1,..., p$ and $j \neq s$.

Given an almost complementary basic feasible solution (w, z, z_0), where w_s and z_s are both nonbasic, an *adjacent almost complementary basic feasible solution* $(\hat{w}, \hat{z}, \hat{z}_0)$ is obtained by introducing either w_s or z_s in the basis such that the pivoting drives a variable other than z_0 from the basis.

From the above definition, it is clear that each almost complementary basic feasible solution has, at most, two adjacent almost complementary basic feasible solutions. If increasing w_s or z_s drives z_0 out of the basis or produces a ray of the set defined in (11.5) and (11.6), then we have less than two adjacent almost complementary basic feasible solutions.

Summary of Lemke's Complementary Pivoting Algorithm

We summarize below a complementary pivoting algorithm credited to Lemke [1968] for solving the linear complementary problem. A similar scheme due to Cottle and Dantzig [1968], known as the *principal pivoting method*, is described in Exercise 11.11. Introducing the artificial variable z_0, the former algorithm moves among adjacent almost complementary basic feasible solutions until either a complementary basic feasible solution is obtained or a direction indicating unboundedness of the region defined by (11.5)–(11.7) is found. As shown later, under certain assumptions on the matrix **M**, the algorithm converges in a finite number of steps, with a complementary basic feasible solution.

Initialization Step If $\mathbf{q} \geq 0$, stop; (w, z) = (q, 0) is a complementary basic feasible solution. Otherwise, display the system defined by (11.5) and (11.6) in a tableau format. Let $-q_s = \max\{-q_i : 1 \leq i \leq p\}$, and update the tableau by pivoting at row s and the z_0 column. Thus, the basic variables z_0 and w_j for $j = 1,..., p$ and $j \neq s$ are nonnegative. Let $y_s = z_s$ and go to the Main Step.

Main Step

1. Let $\mathbf{d}_s$ be the updated column in the current tableau under the variable y_s. If $\mathbf{d}_s \leq 0$, go to Step 4. Otherwise, determine the index r by

the following minimum ratio test, where $\overline{\mathbf{q}}$ is the updated right-hand side column denoting the values of the basic variables:

$$\frac{\overline{q}_r}{d_{rs}} = \min_{1 \le i \le p} \left\{ \frac{\overline{q}_i}{d_{is}} : d_{is} > 0 \right\}.$$

If the basic variable at row r is z_0, go to Step 3. Otherwise, go to Step 2.

2. The basic variable at row r is either w_ℓ or z_ℓ, for some $\ell \neq s$. The variable y_s enters the basis and the tableau is updated by pivoting at row r and the y_s column. If the variable that just left the basis is w_ℓ, then let $y_s = z_\ell$; and if the variable that just left the basis is z_ℓ, then let $y_s = w_\ell$. Go to Step 1.
3. Here y_s enters the basis, and z_0 leaves the basis. Pivot at the y_s column and the z_0 row, producing a complementary basic feasible solution. Stop.
4. Stop with *ray termination.* A ray $R = \{(\mathbf{w}, \mathbf{z}, z_0) + \lambda\mathbf{d} : \lambda \geq 0\}$ is found such that every point in R satisfies (11.5), (11.6), and (11.7). Here, $(\mathbf{w}, \mathbf{z}, z_0)$ is the almost complementary basic feasible solution associated with the last tableau, and $\mathbf{d}$ is an extreme direction of the set defined by (11.5) and (11.6), having a 1 in the row corresponding to y_s, $-\mathbf{d}_s$ in the rows of the current basic variables, and zero everywhere else.

11.1.3 Example (Termination with a Complementary Basic Feasible Solution)

We wish to find a solution to the linear complementary problem defined by

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & -2 \\ 1 & -1 & 2 & -2 \\ 1 & 2 & -2 & 4 \end{bmatrix}, \qquad \mathbf{q} = \begin{bmatrix} 2 \\ 2 \\ -2 \\ -6 \end{bmatrix}.$$

Initialization Step Introduce the artificial variable z_0 and form the following tableau:

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
w_1	1	0	0	0	0	0	1	1	−1	2
w_2	0	1	0	0	0	0	−1	2	−1	2
w_3	0	0	1	0	−1	1	−2	2	−1	−2
w_4	0	0	0	1	−1	−2	2	−4	(−1)	−6

Note that the minimum $\{q_i: 1 \le i \le 4\} = q_4$, so that we pivot at row 4 and the z_0 column. Go to Iteration 1 with $y_s = z_4$.

***Iteration* 1:**

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
w_1	1	0	0	–1	1	2	–1	5	0	8
w_2	0	1	0	–1	1	2	–3	6	0	8
w_3	0	0	1	–1	0	3	–4	(6)	0	4
z_0	0	0	0	–1	1	2	–2	4	1	6

Here $y_s = z_4$ enters the basis. By the minimum ratio test, w_3 leaves the basis; so for the purpose of the next iteration, $y_s = z_3$. We pivot at the w_3 row and the z_4 column, and we go to Iteration 2.

***Iteration* 2:**

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
w_1	1	0	–5/6	–1/6	1	–1/2	(7/3)	0	0	14/3
w_2	0	1	–1	0	1	–1	1	0	0	4
z_4	0	0	1/6	–1/6	0	1/2	–2/3	1	0	2/3
z_0	0	0	–2/3	–1/3	1	0	2/3	0	1	10/3

Here $y_s = z_3$ enters the basis. By the minimum ratio test, w_1 leaves the basis; so for the purpose of the next iteration, $y_s = z_1$. We pivot at the w_1 row and the z_3 column, and we go to Iteration 3.

***Iteration* 3:**

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
z_3	3/7	0	–5/14	–1/14	3/7	–3/14	1	0	0	2
w_2	–3/7	1	–9/14	1/14	4/7	–11/14	0	0	0	2
z_4	2/7	0	–1/14	–3/14	2/7	5/14	0	1	0	2
z_0	–2/7	0	–3/7	–2/7	(5/7)	1/7	0	0	1	2

Here $y_s = z_1$ enters the basis. By the minimum ratio test, z_0 leaves the basis. Pivoting at the z_0 row and the z_1 column gives the complementary basic feasible solution represented by the following tableau:

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
z_3	3/5	0	−1/10	1/10	0	−3/10	1	0	−3/5	4/5
w_3	−1/5	1	−3/10	3/10	0	−9/10	0	0	−4/5	2/5
z_4	2/5	0	1/10	−1/10	0	3/10	0	1	−2/5	6/5
z_1	−2/5	0	−3/5	−2/5	1	1/5	0	0	7/5	14/5

To summarize, the complementary pivoting algorithm produced the point

$$(w_1, w_2, w_3, w_4, z_1, z_2, z_3, z_4) = (0, 2/5, 0, 0, 14/5, 0, 4/5, 6/5),$$

where only one variable from the pair (w_j, z_j) is positive for $j = 1,...,4$.

11.1.4 Example (Ray Termination)

We wish to find a solution to the linear complementary problem defined by

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 2 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 1 \\ 4 \\ -2 \\ -4 \end{bmatrix}.$$

Initialization Step Introduce the artificial variable z_0, leading to the following tableau:

w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
1	0	0	0	0	0	−1	1	−1	1
0	1	0	0	0	0	1	−2	−1	4
0	0	1	0	1	−1	−2	2	−1	−2
0	0	0	1	−1	2	2	−2	−1	−4

Note that $\min\{q_i : 1 \le i \le 4\} = q_4$, so that we pivot at row 4 and the z_0 column. Go to Iteration 1 with $y_s = z_4$.

***Iteration* 1:**

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
w_1	1	0	0	−1	1	−2	−3	3	0	5
w_2	0	1	0	−1	1	−2	−1	0	0	8
w_3	0	0	1	−1	2	−3	−4	(4)	0	2
z_0	0	0	0	−1	1	−2	−2	2	1	4

Here $y_s = z_4$ enters the basis. By the minimum ratio test, w_3 leaves the basis. The tableau is updated by pivoting at the w_3 row and the z_4 column, and we go to Iteration 2 with $y_s = z_3$.

***Iteration* 2:**

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4	z_0	RHS
w_1	1	0	−3/4	−1/4	−1/2	1/4	0	0	0	7/2
w_2	0	1	0	−1	1	−2	−1	0	0	8
z_4	0	0	1/4	−1/4	1/2	−3/4	−1	1	0	1/2
z_0	0	0	−1/2	−1/2	0	−1/2	0	0	1	3

Here $y_s = z_3$ should enter the basis. However, all the entries under the z_3 column are nonpositive, so we stop with ray termination. We have thus found the ray

$$R = \{(\mathbf{w}, \mathbf{z}, z_0) = (7/2, 8, 0, 0, 0, 0, 0, 1/2, 3) + \lambda(0, 1, 0, 0, 0, 0, 1, 1, 0) : \lambda \geq 0\},$$

where every point on the ray satisfies (11.5)–(11.7).

Finite Convergence of the Complementary Pivoting Algorithm

The following lemma shows that the algorithm must stop in a finite number of iterations, either with a complementary basic feasible solution or with ray termination. Under certain conditions on the matrix **M**, the algorithm stops with a complementary basic feasible solution.

11.1.5 Lemma

Suppose that each almost complementary basic feasible solution of the system (11.5)–(11.7) is nondegenerate; that is, each basic variable is positive. Then none of the points generated by the complementary pivoting algorithm is repeated, and furthermore, the algorithm must stop in a finite number of steps.

Proof

Let $(\mathbf{w}, \mathbf{z}, z_0)$ be an almost complementary basic feasible solution, where w_s and z_s are both nonbasic. Then $(\mathbf{w}, \mathbf{z}, z_0)$ has, at most, two adjacent almost complementary basic feasible solutions, one obtained by introducing w_s in the basis and the other obtained by introducing z_s in the basis.* By the nondegeneracy assumption, each of these solutions is distinct from $(\mathbf{w}, \mathbf{z}, z_0)$.

We now show that none of the almost complementary basic feasible solutions generated by the algorithm is repeated. Let $(\mathbf{w}, \mathbf{z}, z_0)_v$ be the point generated at a general iteration v. By contradiction, suppose that $(\mathbf{w}, \mathbf{z}, z_0)_{k+\alpha} = (\mathbf{w}, \mathbf{z}, z_0)_k$ for some positive integers k and α, where $k + \alpha$ is the smallest index for which a repetition is observed. By the nondegeneracy assumption, $\alpha > 1$. Furthermore, by the rules of the algorithm, $\alpha > 2$. But since $(\mathbf{w}, \mathbf{z}, z_0)_{k+\alpha-1}$ is adjacent to $(\mathbf{w}, \mathbf{z}, z_0)_{k+\alpha}$, it is adjacent to $(\mathbf{w}, \mathbf{z}, z_0)_k$. If $k = 1$, and since $(\mathbf{w}, \mathbf{z}, z_0)_k$ has exactly one adjacent almost complementary basic feasible solution, $(\mathbf{w}, \mathbf{z}, z_0)_{k+\alpha-1} = (\mathbf{w}, \mathbf{z}, z_0)_{k+1}$, and hence, a repetition occurs at iteration $k + \alpha - 1$, contradicting our assumption that the first repetition occurs at iteration $k + \alpha$. If $k \geq 2$, then $(\mathbf{w}, \mathbf{z}, z_0)_{k+\alpha-1}$ is adjacent to $(\mathbf{w}, \mathbf{z}, z_0)_k$ and, hence, must be equal to $(\mathbf{w}, \mathbf{z}, z_0)_{k+1}$ or to $(\mathbf{w}, \mathbf{z}, z_0)_{k-1}$. In either case, a repetition occurs at iteration $(\mathbf{w}, \mathbf{z}, z_0)_{k+\alpha-1}$, which contradicts our assumption. Thus, the points generated by the algorithm are distinct.

Since there is only a finite number of almost complementary basic feasible solutions, and since none of them is repeated, the algorithm stops in a finite number of steps with a complementary basic feasible solution or with ray termination. This completes the proof.

To prove the main convergence result specified by Theorem 11.1.8, Lemma 11.1.6 and Definition 11.1.7 are needed. The lemma gives certain implications of ray termination, and the definition introduces the concept of a copositive-plus matrix.

11.1.6 Lemma

Suppose that each almost complementary basic feasible solution of the system defined by (11.5)–(11.7) is nondegenerate. Suppose that the complementary pivoting algorithm is used to solve this system, and further, suppose that ray termination occurs. In particular, assume that at termination we have the almost complementary basic feasible solution $(\overline{\mathbf{w}}, \overline{\mathbf{z}}, \overline{z}_0)$ and the extreme direction $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0)$, giving the ray $R = \{(\overline{\mathbf{w}}, \overline{\mathbf{z}}, \overline{z}_0) + \lambda(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0) : \lambda \geq 0\}$. Then

1. $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0) \neq (0, 0, 0)$, $(\hat{\mathbf{w}}, \hat{\mathbf{z}}) \geq \mathbf{0}$, $\hat{z}_0 \geq 0$.

---

* Note that $(\mathbf{w}, \mathbf{z}, z_0)$ may have less than two adjacent almost complementary basic feasible solutions. In this case the column under w_s or z_s is ≤ 0, or else introducing w_s or z_s in the basis drives z_0 out of the basis, thus producing a complementary basic feasible solution.

2. $\hat{\mathbf{w}} - \mathbf{M}\hat{\mathbf{z}} - \mathbf{e}\hat{z}_0 = \mathbf{0}$.
3. $\overline{\mathbf{w}}^t\overline{\mathbf{z}} = \overline{\mathbf{w}}^t\hat{\mathbf{z}} = \hat{\mathbf{w}}^t\overline{\mathbf{z}} = \hat{\mathbf{w}}^t\hat{\mathbf{z}} = 0$.
4. $\hat{\mathbf{z}} \neq \mathbf{0}$.
5. $\hat{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} = -\mathbf{e}^t\hat{\mathbf{z}}\hat{z}_0 \leq 0$.

Proof

Since $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0)$ is an extreme direction of the set defined by (11.5) and (11.6), Parts 1 and 2 are immediate by Theorem 2.6.6. Recall that every point on the ray R satisfies (11.7), so that $0 = (\overline{\mathbf{w}} + \lambda\hat{\mathbf{w}})^t(\overline{\mathbf{z}} + \lambda\hat{\mathbf{z}})$ for each $\lambda \geq 0$. This, together with the nonnegativity of $\overline{\mathbf{w}}$, $\hat{\mathbf{w}}$, $\overline{\mathbf{z}}$, and $\hat{\mathbf{z}}$, implies that

$$\overline{\mathbf{w}}^t\overline{\mathbf{z}} = \overline{\mathbf{w}}^t\hat{\mathbf{z}} = \hat{\mathbf{w}}^t\overline{\mathbf{z}} = \hat{\mathbf{w}}^t\hat{\mathbf{z}} = 0. \tag{11.8}$$

Therefore, Part 3 holds true.

We now show that $\hat{\mathbf{z}} \neq \mathbf{0}$. By contradiction, suppose that $\hat{\mathbf{z}} = \mathbf{0}$. Note that $\hat{z}_0 > 0$, because otherwise, $\hat{z}_0 = 0$; and from Part 2, we get $\hat{\mathbf{w}} = \mathbf{0}$, contradicting the fact that $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0) \neq (0, 0, 0)$. Thus, $\hat{z}_0 > 0$ and $\hat{\mathbf{w}} = \mathbf{e}\hat{z}_0$.

We have proved that if $\hat{\mathbf{z}} = \mathbf{0}$, then $\hat{z}_0 > 0$ and $\hat{\mathbf{w}} = \mathbf{e}\hat{z}_0$. From (11.8) we get $0 = \hat{\mathbf{w}}^t\overline{\mathbf{z}}$. Thus, $\mathbf{e}^t\overline{\mathbf{z}} = 0$; and since $\overline{\mathbf{z}} \geq \mathbf{0}$, we get $\overline{\mathbf{z}} = \mathbf{0}$. By the nondegeneracy assumption, every component of $\overline{\mathbf{z}}$ is nonbasic. Furthermore, $\overline{z}_0$ is basic, and we must have exactly $p - 1$ basic components of $\overline{\mathbf{w}}$. In particular, since $\overline{\mathbf{w}} - \mathbf{M}\overline{\mathbf{z}} - \mathbf{e}\overline{z}_0 = \mathbf{q}$ and since $\overline{\mathbf{z}} = \mathbf{0}$, we get $\overline{z}_0 = \max\{-q_i : 1 \leq i \leq p\}$. This shows that the almost complementary basic feasible solution $(\overline{\mathbf{w}}, \overline{\mathbf{z}}, \overline{z}_0)$ is the starting solution, which is impossible by Lemma 11.1.5. Therefore, $\hat{\mathbf{z}} \neq \mathbf{0}$ and Part 4 holds true. Multiplying $\hat{\mathbf{w}} - \mathbf{M}\hat{\mathbf{z}} - \mathbf{e}\hat{z}_0 = \mathbf{0}$ by $\hat{\mathbf{z}}^t$, and noting from (11.8) that $\hat{\mathbf{z}}^t\hat{\mathbf{w}} = 0$, we have $\hat{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} = -\hat{\mathbf{z}}^t\mathbf{e}\hat{z}_0 \leq 0$ and Part 5 follows. This completes the proof.

11.1.7 Definition

Let $\mathbf{M}$ be a $p \times p$ matrix. Then $\mathbf{M}$ is said to be *copositive* if $\mathbf{z}^t\mathbf{M}\mathbf{z} \geq 0$ for each $\mathbf{z} \geq \mathbf{0}$. Furthermore, $\mathbf{M}$ is said to be *copositive-plus* if it is copositive and if $\mathbf{z} \geq 0$ and $\mathbf{z}^t\mathbf{M}\mathbf{z} = 0$ imply that $(\mathbf{M} + \mathbf{M}^t)\mathbf{z} = \mathbf{0}$.

Theorem 11.1.8 shows that if the system defined by (11.1) and (11.2) is consistent, and if the matrix $\mathbf{M}$ is copositive plus, then the complementary pivoting algorithm will produce a complementary basic feasible solution in a finite number of steps.

11.1.8 Theorem

Suppose that each almost complementary basic feasible solution to the system defined by (11.5)–(11.7) is nondegenerate, and suppose that **M** is copositive plus. Then the complementary pivoting algorithm stops in a finite number of steps. In particular, if the system defined by (11.1) and (11.2) is consistent, the algorithm stops with a complementary basic feasible solution to the system defined by (11.1)–(11.3). On the other hand, if the system defined in (11.1) and (11.2) is inconsistent, the algorithm stops with ray termination.

Proof

By Lemma 11.1.5, the complementary pivoting algorithm stops in a finite number of steps. Now suppose that the algorithm stops with ray termination. In particular, suppose that $(\overline{\mathbf{w}}, \overline{\mathbf{z}}, \overline{z}_0)$ is the almost complementary basic feasible solution and $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0)$ is the extreme direction associated with the final tableau. By Lemma 11.1.6,

$$\hat{\mathbf{z}} \geq \mathbf{0}, \qquad \hat{\mathbf{z}} \neq \mathbf{0}, \qquad \text{and} \qquad \hat{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} = -\mathbf{e}^t\hat{\mathbf{z}}\hat{z}_0 \leq 0. \tag{11.9}$$

But since **M** is copositive plus, $\hat{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} \geq 0$. From (11.9) it follows that $0 = \hat{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} = -\mathbf{e}^t\hat{\mathbf{z}}\hat{z}_0$. Since $\hat{\mathbf{z}} \neq \mathbf{0}$, $\hat{z}_0 = 0$. But since $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0)$ is a direction of the set defined by (11.5) and (11.6), $\hat{\mathbf{w}} - \mathbf{M}\hat{\mathbf{z}} - \mathbf{e}\hat{z}_0 = \mathbf{0}$, and hence,

$$\hat{\mathbf{w}} = \mathbf{M}\hat{\mathbf{z}}. \tag{11.10}$$

We now show that $\mathbf{q}^t\hat{\mathbf{z}} < 0$. Since $\hat{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} = 0$ and **M** is copositive plus $(\mathbf{M} + \mathbf{M}^t)\hat{\mathbf{z}} = \mathbf{0}$. This, together with Part 3 of Lemma 11.1.6 and the fact that $\overline{\mathbf{w}} = \mathbf{q} + \mathbf{M}\overline{\mathbf{z}} + \mathbf{e}\overline{z}_0$, implies that

$$0 = \overline{\mathbf{w}}^t\hat{\mathbf{z}} = (\mathbf{q} + \mathbf{M}\overline{\mathbf{z}} + \mathbf{e}\overline{z}_0)^t\hat{\mathbf{z}} = \mathbf{q}^t\hat{\mathbf{z}} - \overline{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} + \overline{z}_0\mathbf{e}^t\hat{\mathbf{z}}. \tag{11.11}$$

From (11.10), $\mathbf{M}\hat{\mathbf{z}} = \hat{\mathbf{w}}$, and hence from Part 3 of Lemma 11.1.6, it follows that $\overline{\mathbf{z}}^t\mathbf{M}\hat{\mathbf{z}} = 0$. Furthermore, $\overline{z}_0 > 0$ and $\mathbf{e}^t\hat{\mathbf{z}} > 0$ by (11.9). Substituting in (11.11), it follows that $\mathbf{q}^t\hat{\mathbf{z}} < 0$.

To summarize, we have shown that $\mathbf{M}\hat{\mathbf{z}} = \hat{\mathbf{w}} \geq \mathbf{0}$. Since $(\mathbf{M} + \mathbf{M}^t)\hat{\mathbf{z}} = \mathbf{0}$, we get $\mathbf{M}^t\hat{\mathbf{z}} = -\mathbf{M}\hat{\mathbf{z}} \leq \mathbf{0}$, $-\mathbf{I}\hat{\mathbf{z}} \leq \mathbf{0}$, and $\mathbf{q}^t\hat{\mathbf{z}} < 0$. Thus, the system $\mathbf{M}^t\mathbf{y} \leq \mathbf{0}$, $-\mathbf{I}\mathbf{y} \leq \mathbf{0}$, and $\mathbf{q}^t\mathbf{y} < 0$ has a solution, say, $\mathbf{y} = \hat{\mathbf{z}}$. By Theorem 2.4.5 it follows that the system $\mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q}$, $\mathbf{w} \geq \mathbf{0}$, $\mathbf{z} \geq \mathbf{0}$ has no solution.

Now if the system defined by (11.1) and (11.2) is consistent, then the algorithm must stop with a complementary basic feasible solution, because otherwise, the algorithm would stop with ray termination, which, as we showed above, is possible only if the system (11.1) and (11.2) is inconsistent. If the system defined by (11.1) and (11.2) is inconsistent, then the algorithm obviously could not stop with a complementary basic feasible solution and, hence, must stop with ray termination. This completes the proof.

Corollary

If $\mathbf{M}$ has nonnegative entries, with positive diagonal elements, then the complementary pivoting algorithm stops in a finite number of steps with a complementary basic feasible solution.

Proof

First, note that by the stated assumption on $\mathbf{M}$, the system $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $(\mathbf{w}, \mathbf{z}) \geq \mathbf{0}$ has a solution, say, by choosing $\mathbf{z}$ sufficiently large so that $\mathbf{w} = \mathbf{Mz} + \mathbf{q} \geq \mathbf{0}$. The result then follows from the theorem by noting that $\mathbf{M}$ is copositive plus.

When $\mathbf{M}$ is a general $p \times p$ matrix, the complementary pivoting algorithm might fail to solve the linear complementary problem. In such a case, we can resort to using the aforementioned mixed-integer zero–one bilinear programming formulation of this problem and apply a suitable *reformulation–linearization technique* (RLT), as discussed in Exercises 11.4, 11.6, and 11.27 (See also the Notes and References section for further details.)

11.2 Convex and Nonconvex Quadratic Programming: Global Optimization Approaches

In this section we consider the following quadratic programming problem:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned}$$

where $\mathbf{c}$ is an n-vector, $\mathbf{b}$ is an m-vector, $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{H}$ is an $n \times n$ symmetric matrix. (Note that a more general set of linear constraints can be cast in this format through standard linear transformations. In particular, if the constraints were of the form $\mathbf{A}'\mathbf{x}' = \mathbf{b}'$, $\mathbf{x}' \geq \mathbf{0}$, then the above type of constraints might be an equivalent representation of this region in some nonbasic variable space, using the partitioning scheme described in Section 10.6.)

Observe that the above quadratic program represents a special class of nonlinear programming problems in which the objective function is quadratic and the constraints are linear. In this section we show that the KKT conditions of a quadratic programming problem reduce to a linear complementary problem. Thus, the complementary pivoting algorithm described in Section 11.1 can be used for solving quadratic programming problems.

Several other special procedures for solving quadratic programming problems are discussed in the exercises at the end of the chapter. In particular, Exercise 11.18 shows that if the quadratic programming problem is of the form to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to only equality constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix of rank m and $\mathbf{H}$ is positive definite on $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}$, the unique solution to this problem is obtainable via the solution of the linear KKT system of equations, typically done using an LU factorization approach (see Appendix A.2). If inequality constraints are present, we can use the reduced gradient method as described in Chapter 10, or, as is most popularly done, we can adopt an *active set* strategy as follows. Here, given a feasible solution, an equality-constrained quadratic programming problem is solved to find a correction direction over the nullspace of the active constraints; and either optimality is verified or the set of designated active constraints (to be held as equalities) is modified, at the current solution itself or at a revised solution, and the process is repeated. Exercises 11.19 and 11.28 describe such strategies. For convex quadratic programming problems, we can also extend the primal-dual path-following algorithm described in Chapter 9, using the barrier penalty function algorithm along with Newton's method, to derive a polynomial-time algorithm. The Notes and References section directs the reader to this and to other interior point approaches for convex quadratic programming problems.

When the Hessian of the gradient function is not positive semidefinite, however, finding a global minimum for the underlying quadratic programming problem becomes a difficult task. In fact, quadratic (minimization) problems with even a single negative eigenvalue for the Hessian are known to be NP-hard. Assuming that an optimum exists, Exercise 11.5 shows how a general quadratic program can be posed as a linear program with complementarity constraints, for which some of the zero-one linearization approaches suggested in the foregoing section and in Exercise 11.4 can be employed. Another approach that has been demonstrated to be particularly effective for this class of problems is to apply a *reformulation-linearization/convexification technique* (RLT), as discussed in the sequel (also see the Notes and References section for further extensions). This approach generates a linear programming relaxation of the original quadratic problem through a *reformulation step* that generates additional restrictions via suitable pairwise products of constraints, followed by a *linearization step* that linearizes the resulting problem by substituting a new variable w_{ij} in place of each nonlinear quadratic term $x_i x_j$, $\forall\ 1 \le i \le j \le n$. This relaxation yields tight lower bounds on the original problem and possesses certain desirable properties (see Lemmas 11.2.5 and 11.2.6), which enables embedding it within a specially designed branch-and-bound algorithm that can be proven to recover a global optimal solution to the underlying quadratic program. Moreover, this approach can be extended to solve more general *polynomial programming problems* having polynomial objective and constraint functions to global optimality (see Exercise 11.27 and the Notes and References section).

We now proceed to consider the solution of quadratic programs through linear complementary problems.

Karush–Kuhn–Tucker System

Consider the above quadratic programming problem. Denoting the Lagrangian multiplier vectors of the constraints $\mathbf{Ax} \geq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ by u and v, respectively, and denoting the vector of slack variables by y, the KKT conditions for this problem can be written as follows:

$$\begin{aligned} \mathbf{Ax}+\mathbf{y} &= \mathbf{b} \\ -\mathbf{Hx}-\mathbf{A}^t\mathbf{u}+\mathbf{v} &= \mathbf{c} \\ \mathbf{x}^t\mathbf{v}=0, \quad \mathbf{u}^t\mathbf{y} &= 0 \\ \mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v} &\geq \mathbf{0}. \end{aligned}$$

Now, letting

$$\mathbf{M}=\begin{bmatrix} \mathbf{0} & -\mathbf{A} \\ \mathbf{A}^t & \mathbf{H} \end{bmatrix}, \quad \mathbf{q}=\begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix}, \quad \mathbf{w}=\begin{pmatrix} \mathbf{y} \\ \mathbf{v} \end{pmatrix}, \quad \text{and} \quad \mathbf{z}=\begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix},$$

we can rewrite the KKT conditions as the linear complementary problem $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $\mathbf{w}^t\mathbf{z} = 0$, $(\mathbf{w}, \mathbf{z}) \geq \mathbf{0}$. Thus, the complementary pivoting algorithm discussed in Section 11.1 can be used to find a KKT point of the quadratic programming problem.

11.2.1 Example (Finite Optimal Solution)

Consider the following quadratic programming problem:

$$\begin{aligned} \text{Minimize } & -2x_1 - 6x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{subject to } & \quad x_1 + x_2 \leq 2 \\ & -x_1 + 2x_2 \leq 2 \\ & \quad x_1, \quad x_2 \geq 0. \end{aligned}$$

Note that

$$\mathbf{A}=\begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{H}=\begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}, \quad \mathbf{b}=\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{c}=\begin{bmatrix} -2 \\ -6 \end{bmatrix}.$$

Denote the vector of slacks by y and the Lagrangian multiplier vectors for the constraints $\mathbf{Ax} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ by $\mathbf{u}$ and v, respectively. Let

$$\mathbf{M}=\begin{bmatrix} \mathbf{0} & -\mathbf{A} \\ \mathbf{A}^t & \mathbf{H} \end{bmatrix}, \quad \mathbf{q}=\begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix}, \quad \mathbf{w}=\begin{bmatrix} \mathbf{y} \\ \mathbf{v} \end{bmatrix}, \quad \text{and} \quad \mathbf{z}=\begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix}.$$

Then the KKT conditions reduce to finding a solution to the system $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $\mathbf{w}^t\mathbf{z} = 0$, and $(\mathbf{w}, \mathbf{z}) \geq \mathbf{0}$, where

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & -2 \\ 1 & -1 & 2 & -2 \\ 1 & 2 & -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} 2 \\ 2 \\ -2 \\ -6 \end{bmatrix}.$$

The problem of finding a complementary basic feasible solution to the above system was solved in Example 11.1.3, producing the KKT point $(x_1, x_2) = (z_3, z_4) = (4/5, 6/5)$. Reviewing Example 11.1.3, note that the complementary pivoting algorithm started from the point (0, 0), then moved to the point (0, 2/3), then to the point (2, 2), and finally, to the KKT point (4/5, 6/5). Since **H** is positive definite, the objective function is convex, so the KKT point (4/5, 6/5) is indeed optimal. The path taken by the complementary pivoting algorithm to produce the optimal solution is shown in Figure 11.1.

11.2.2 Example (Unbounded Optimal Solution)

Consider the following quadratic programming problem:

$$\begin{aligned} \text{Minimize } & -2x_1 - 4x_2 + x_1^2 - 2x_1x_2 + x_2^2 \\ \text{subject to } & -x_1 + x_2 \le 1 \\ & x_1 - 2x_2 \le 4 \\ & x_1, \quad x_2 \ge 0. \end{aligned}$$

Note that

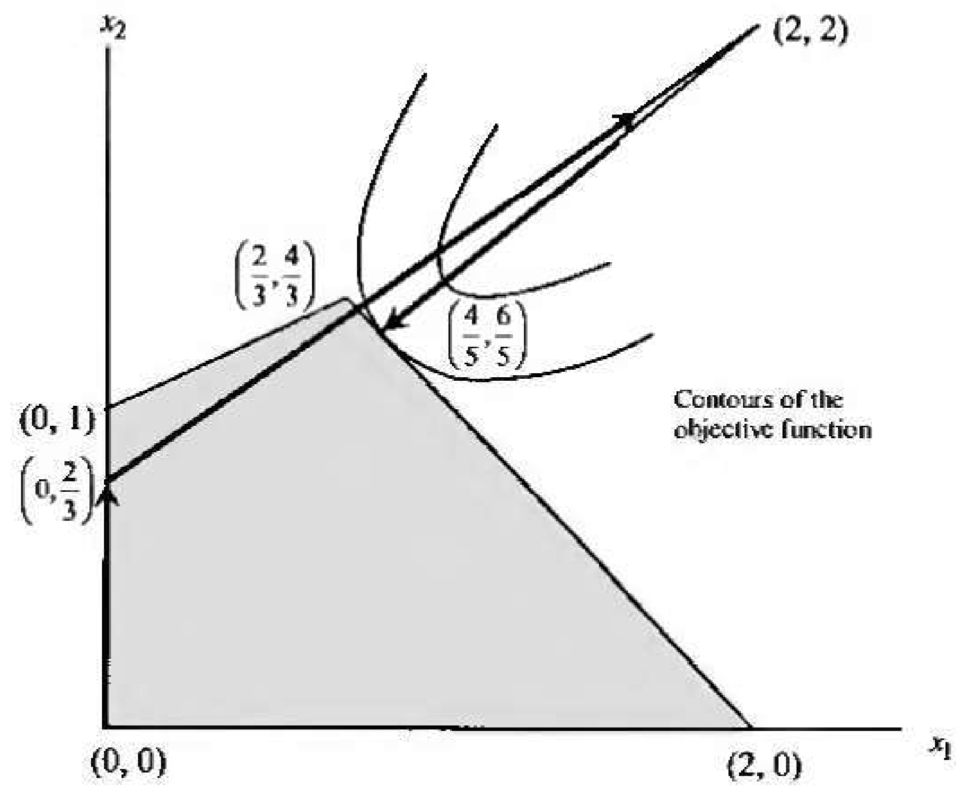


Figure 11.1 Points generated by the complementary pivoting algorithm.

$$\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}.$$

Denote the vector of slacks by $\mathbf{y}$ and the Lagrangian multiplier vectors for the constraints $\mathbf{Ax} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$ by u and v, respectively. Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & -\mathbf{A} \\ \mathbf{A}^t & \mathbf{H} \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix}, \quad \mathbf{w} = \begin{pmatrix} \mathbf{y} \\ \mathbf{v} \end{pmatrix}, \quad \text{and} \quad \mathbf{z} = \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix}.$$

Then, solving the KKT conditions reduces to finding a solution to the system $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $\mathbf{w}^t\mathbf{z} = 0$, and $(\mathbf{w}, \mathbf{z}) \ge \mathbf{0}$, where

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 2 \end{bmatrix} \text{ and } \mathbf{q} = \begin{bmatrix} 1 \\ 4 \\ -2 \\ -4 \end{bmatrix}.$$

The problem of finding a complementary basic feasible solution to the above system was solved in Example 11.1.4. As shown in that example, the complementary pivoting algorithm stopped with ray termination and was unable to produce a complementary basic feasible solution. The reason for this is that the optimal solution is unbounded along the ray R produced by the algorithm. Mapped in the (x_1, x_2) space, the ray $R = \{(0, 1/2) + \lambda(1, 1) : \lambda \ge 0\}$ leads to an unbounded optimal solution, as shown in Figure 11.2.

Convergence Analysis of the Quadratic Programming Complementary Pivoting Algorithm

In Section 11.1 we showed that under nondegeneracy, the complementary pivoting algorithm stops in a finite number of steps with either a complementary basic feasible solution or a ray termination. We also showed that if the matrix $\mathbf{M}$ associated with the linear complementary problem is copositive plus, and the linear constraints are consistent, the algorithm produces a complementary basic feasible solution. Theorem 11.2.3 gives some sufficient conditions for the matrix $\mathbf{M}$ associated with the quadratic problem to be copositive plus. Following this, Theorem 11.2.4 gives several conditions under which the complementary pivoting algorithm produces a KKT point and shows that ray termination is only possible if the quadratic programming problem has an unbounded optimal solution.

11.2.3 Theorem

Let $\mathbf{A}$ be an $m \times n$ matrix, and let $\mathbf{H}$ be an $n \times n$ symmetric matrix. If $\mathbf{y}^t\mathbf{H}\mathbf{y} \ge 0$ for each $\mathbf{y} \ge \mathbf{0}$, then the matrix

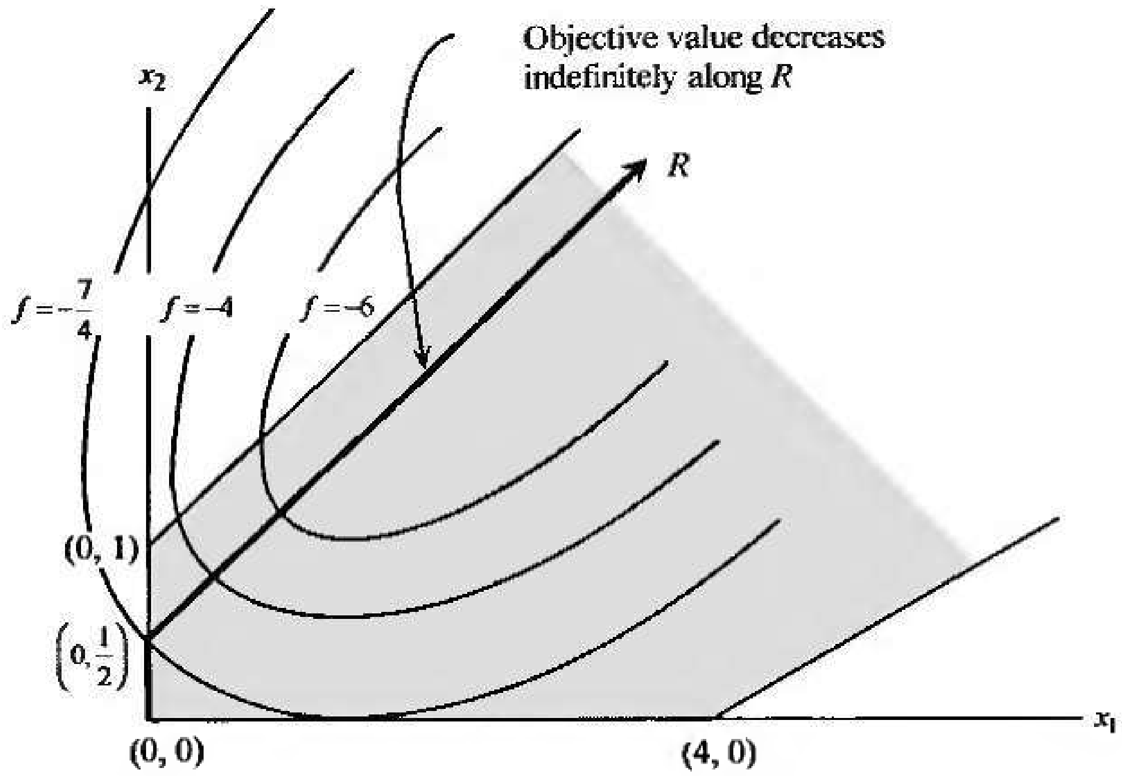


Figure 11.2 Unbounded optimal solution and ray termination.

$$\mathbf{M}=\begin{bmatrix} \mathbf{0} & -\mathbf{A} \\ \mathbf{A}^t & \mathbf{H} \end{bmatrix}$$

is copositive. In addition, if $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$ imply that $\mathbf{H}\mathbf{y} = \mathbf{0}$, then $\mathbf{M}$ is copositive plus.

Proof

First, we show that $\mathbf{M}$ is copositive. Let $\mathbf{z}^t = (\mathbf{x}^t, \mathbf{y}^t) \geq \mathbf{0}$. Then

$$\mathbf{z}^t\mathbf{M}\mathbf{z} = (\mathbf{x}^t, \mathbf{y}^t)\begin{bmatrix} \mathbf{0} & -\mathbf{A} \\ \mathbf{A}^t & \mathbf{H} \end{bmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{y}^t\mathbf{H}\mathbf{y}. \tag{11.12}$$

By assumption, $\mathbf{y}^t\mathbf{H}\mathbf{y} \geq 0$, and hence $\mathbf{H}$ is copositive. To show that $\mathbf{M}$ is copositive plus, suppose that $\mathbf{z} \geq \mathbf{0}$ and $\mathbf{z}^t\mathbf{M}\mathbf{z} = 0$. It suffices to show that $(\mathbf{M} + \mathbf{M}^t)\mathbf{z} = \mathbf{0}$. But

$$\mathbf{M}+\mathbf{M}^t=\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{H} \end{bmatrix}$$

and hence

$$(\mathbf{M}+\mathbf{M}^t)\mathbf{z}=\begin{bmatrix}\mathbf{0}\\ 2\mathbf{H}\mathbf{y}\end{bmatrix}.$$

Since $\mathbf{z}^t\mathbf{M}\mathbf{z} = 0$, we get $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$ by (11.12). By assumption, since $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$, we get $\mathbf{H}\mathbf{y} = \mathbf{0}$, and hence $(\mathbf{M} + \mathbf{M}^t)\mathbf{z} = \mathbf{0}$, so that $\mathbf{M}$ is copositive plus. This completes the proof.

Corollary 1

If $\mathbf{H}$ is positive semidefinite, then $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$ implies that $\mathbf{H}\mathbf{y} = \mathbf{0}$, so $\mathbf{M}$ is copositive plus.

Proof

It suffices to show that $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$ implies that $\mathbf{H}\mathbf{y} = \mathbf{0}$. Let $\mathbf{H}\mathbf{y} = \mathbf{d}$, and noting that $\mathbf{H}$ is positive semidefinite, we get

$$0 \leq (\mathbf{y}^t - \lambda\mathbf{d}^t)\mathbf{H}(\mathbf{y}-\lambda\mathbf{d}) = \mathbf{y}^t\mathbf{H}\mathbf{y} + \lambda^2\mathbf{d}^t\mathbf{H}\mathbf{d} - 2\lambda\|\mathbf{d}\|^2.$$

Since $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$, dividing the above inequality by λ and letting $\lambda \to 0^+$, it follows that $\mathbf{0} = \mathbf{d} = \mathbf{H}\mathbf{y}$.

Corollary 2

If $\mathbf{H}$ has nonnegative entries, then $\mathbf{M}$ is copositive. Furthermore, if $\mathbf{H}$ has nonnegative elements with positive diagonal elements, then $\mathbf{M}$ is copositive plus.

Proof

If $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}^t\mathbf{H}\mathbf{y} = 0$, then $\mathbf{y} = \mathbf{0}$, and hence $\mathbf{H}\mathbf{y} = \mathbf{0}$. By the theorem, $\mathbf{M}$ is copositive plus.

11.2.4 Theorem

Consider the problem to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$. Suppose that the feasible region is not empty. Further, suppose that the complementary pivoting algorithm described in Section 11.1 is used in an attempt to find a solution to the KKT system $\mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q}$, $(\mathbf{w}, \mathbf{z}) \geq \mathbf{0}$, $\mathbf{w}^t\mathbf{z} = 0$, where

$$\mathbf{M}=\begin{bmatrix}\mathbf{0} & -\mathbf{A}\\ \mathbf{A}^t & \mathbf{H}\end{bmatrix}, \qquad \mathbf{q}=\begin{pmatrix}\mathbf{b}\\ \mathbf{c}\end{pmatrix}, \qquad \mathbf{w}=\begin{pmatrix}\mathbf{y}\\ \mathbf{v}\end{pmatrix}, \qquad \mathbf{z}=\begin{pmatrix}\mathbf{u}\\ \mathbf{x}\end{pmatrix},$$

$\mathbf{y}$ is the vector of slack variables, and $\mathbf{u}$ and $\mathbf{v}$ are the Lagrangian multiplier vectors associated with the constraints $\mathbf{Ax} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$, respectively. In the absence of degeneracy, under any of the following conditions, the algorithm stops in a finite number of iterations with a KKT point:

1. $\mathbf{H}$ is positive semidefinite and $\mathbf{c} \ge \mathbf{0}$.
2. $\mathbf{H}$ is positive definite.
3. $\mathbf{H}$ has nonnegative elements with positive diagonal elements.

Moreover, if $\mathbf{H}$ is positive semidefinite, then ray termination implies that the optimal solution is unbounded.

Proof

Assume that $\mathbf{H} = \mathbf{H}^t$, because otherwise, $\mathbf{H}$ could be replaced by $(1/2)(\mathbf{H} + \mathbf{H}^t)$. From Lemma 11.1.5, the complementary pivoting algorithm stops in a finite number of iterations with either a KKT point or a ray termination. If $\mathbf{H}$ is positive semidefinite or positive definite, or has nonnegative elements with positive diagonal elements, then, by Corollaries 1 and 2 to Theorem 11.2.3, $\mathbf{M}$ is copositive plus.

Now suppose that ray termination occurs. By Theorem 11.1.8, since $\mathbf{M}$ is copositive plus, ray termination is possible only if the following system has no solution:

$$\begin{aligned} \mathbf{Ax}+\mathbf{y} &= \mathbf{b} \\ -\mathbf{Hx}-\mathbf{A}^t\mathbf{u}+\mathbf{v} &= \mathbf{c} \\ \mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} &\ge \mathbf{0}. \end{aligned}$$

By Theorem 2.4.5, the following system must have a solution $(\mathbf{d}, \mathbf{f})$:

$$\mathbf{Ad} \le \mathbf{0} \tag{11.13a}$$

$$\mathbf{A}^t\mathbf{f} - \mathbf{Hd} \ge \mathbf{0} \tag{11.13b}$$

$$\mathbf{f} \ge \mathbf{0} \tag{11.13c}$$

$$\mathbf{d} \ge \mathbf{0} \tag{11.13d}$$

$$\mathbf{b}^t\mathbf{f} + \mathbf{c}^t\mathbf{d} < 0. \tag{11.13e}$$

Multiplying (11.13b) by $\mathbf{d}^t \ge \mathbf{0}$ and noting that $\mathbf{f} \ge \mathbf{0}$ and $\mathbf{Ad} \le \mathbf{0}$, it follows that

$$0 \le \mathbf{d}^t\mathbf{A}^t\mathbf{f} - \mathbf{d}^t\mathbf{Hd} \le 0 - \mathbf{d}^t\mathbf{Hd} = -\mathbf{d}^t\mathbf{Hd}. \tag{11.14a}$$

By assumption, there exist $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ such that $\mathbf{A}\hat{\mathbf{x}} + \hat{\mathbf{y}} = \mathbf{b}$, $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \ge \mathbf{0}$. Substituting this for $\mathbf{b}$ in (11.13e) and noting (11.13b) and that $(\mathbf{f}, \hat{\mathbf{x}}, \hat{\mathbf{y}}) \ge \mathbf{0}$, we get

$$\mathbf{0} > \mathbf{c}^t\mathbf{d} + \mathbf{b}^t\mathbf{f} = \mathbf{c}^t\mathbf{d} + (\hat{\mathbf{y}} + \mathbf{A}\hat{\mathbf{x}})^t\mathbf{f} \ge \mathbf{c}^t\mathbf{d} + \hat{\mathbf{x}}^t\mathbf{A}^t\mathbf{f} \ge \mathbf{c}^t\mathbf{d} + \hat{\mathbf{x}}^t\mathbf{Hd}. \tag{11.14b}$$

Now, suppose that **H** is positive semidefinite. By (11.14a) it follows that $\mathbf{d}^t\mathbf{H}\mathbf{d} = 0$, and by Corollary 1 to Theorem 11.2.3 it follows that $\mathbf{H}\mathbf{d} = \mathbf{0}$. By (11.14b) we have $\mathbf{c}^t\mathbf{d} < 0$. Since $\mathbf{A}\mathbf{d} \le \mathbf{0}$ and $\mathbf{d} \ge \mathbf{0}$, d is a direction of the feasible region, so that $\hat{\mathbf{x}} + \lambda\mathbf{d}$ is feasible for all $\lambda \ge 0$. Now consider $f(\hat{\mathbf{x}}+\lambda\mathbf{d})$, where $f(\mathbf{x}) = \mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$. Since $\mathbf{H}\mathbf{d} = \mathbf{0}$, we get

$$f(\hat{\mathbf{x}}+\lambda\mathbf{d}) = f(\hat{\mathbf{x}}) + \lambda(\mathbf{c}^t + \hat{\mathbf{x}}^t\mathbf{H})\mathbf{d} + \frac{1}{2}\lambda^2\mathbf{d}^t\mathbf{H}\mathbf{d} = f(\hat{\mathbf{x}}) + \lambda\mathbf{c}^t\mathbf{d}.$$

Since $\mathbf{c}^t\mathbf{d} < 0$, $f(\hat{\mathbf{x}}+\lambda\mathbf{d})$ approaches $-\infty$ by choosing λ arbitrarily large; thus we have an unbounded optimal solution.

To complete the proof, we now show that ray termination is not possible under Condition 1, 2, or 3 of the theorem. On the contrary, suppose that ray termination occurs under any of these conditions. From (11.14a), $\mathbf{d}^t\mathbf{H}\mathbf{d} \le 0$. Under Condition 2 or 3, $\mathbf{d} = \mathbf{0}$, which is impossible in view of (11.14b). If Condition 1 holds true, on the other hand, then $\mathbf{H}\mathbf{d} = \mathbf{0}$ as above. This, together with (11.13d) and the assumption that $\mathbf{c} \ge \mathbf{0}$ contradicts (11.14b).

To summarize, we have shown that if **H** is positive semidefinite and the algorithm stops with ray termination, the optimal solution is unbounded. Furthermore, ray termination is impossible under Condition 1, 2, or 3, so the algorithm must produce a KKT point under any of these conditions. This completes the proof.

Global Optimization Approach for Nonconvex Quadratic Programs

Consider a *nonconvex quadratic programming problem* stated in the following form:

$$\text{NQP: Minimize } \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x}$$

$$\text{subject to } \mathbf{A}\mathbf{x} \le \mathbf{b}$$

$$\mathbf{x} \in \Omega \equiv \{\mathbf{x} : \ell_j \le x_j \le u_j \text{ for } j = 1,\dots, n\},$$

where **H** is $n \times n$, symmetric, but not necessarily positive semidefinite, **A** is an $m \times n$ matrix, and where the *hyperrectangle* Ω defines finite lower and upper bounds on the variables, with $\ell_j < u_j$, $\forall j = 1,\dots, n$. Moreoever, let us assume that the feasible region is nonempty, so that by Weierstrass's theorem, an optimum exists. For the sake of convenience, letting $h_{k\ell}$ denote then the (k, ℓ)th element of **H** and denoting the collection of $m + 2n$ inequalities $\mathbf{A}\mathbf{x} \le \mathbf{b}$, and $\ell_j \le x_j \le u_j$ for $j = 1,\dots, n$ jointly by $\mathbf{G}_i\mathbf{x} \equiv \sum_{k=1}^{n} G_{ik}x_k \le g_i$, for $i = 1,\dots, \bar{m} \equiv m + 2n$, let us rewrite Problem NQP as follows:

$$\text{NQP: Minimize } \sum_{k=1}^{n} c_k x_k + \frac{1}{2}\sum_{k=1}^{n}\sum_{\ell=1}^{n} h_{k\ell} x_k x_\ell \qquad (11.15)$$
$$\text{subject to } \mathbf{G}_i \mathbf{x} \le g_i \quad \text{for } i = 1, \ldots, \bar{m}.$$

We shall now describe a specialized rudimentary application of a *reformulation-linearization/convexification technique* (RLT) to solve this nonconvex quadratic program NQP to global optimality. (Several refinements to this basic approach will be discussed subsequently.) As its name suggests, the RLT process operates in two phases: a reformulation phase and a linearization (or convexification) phase. In the *reformulation phase*, we replace the constraints (11.15) with a pairwise product of these restrictions, namely, with

$$(g_i - \mathbf{G}_i\mathbf{x})(g_j - \mathbf{G}_j\mathbf{x}) \ge 0 \qquad \text{for } 1 \le i \le j \le \bar{m}.$$

Following this, we apply the *linearization phase* in which each distinct quadratic term $x_k x_\ell$, for $1 \le k \le \ell \le n$, is replaced by a single *new RLT variable* $w_{k\ell}$. In other words, we simply substitute

$$w_{k\ell} = x_k x_\ell \qquad \text{for } 1 \le k \le \ell \le n. \qquad (11.16)$$

For example, given some two defining inequalities $2x_1 + 3x_2 \le 6$ and $x_1 - 2x_2 \le 2$, we would linearize the product $(6 - 2x_1 - 3x_2)\ (2 - x_1 + 2x_2) \ge 0$ using the substitution $w_{11} = x_1^2$, $w_{22} = x_2^2$, and $w_{12} = x_1x_2$ to obtain $2w_{11} - 6w_{22} - w_{12} - 10x_1 + 6x_2 + 12 \ge 0$. Let us represent the latter linearized inequality more transparently as $[(6 - 2x_1 - 3x_2)(2 - x_1 + 2x_2)]_L \ge 0$. In particular, under (11.16), we have

$$\{[(g_i - \mathbf{G}_i\mathbf{x})(g_j - \mathbf{G}_j\mathbf{x})]_L \ge 0\}$$
$$\equiv \left\{ g_i g_j - g_i \mathbf{G}_j \mathbf{x} - g_j \mathbf{G}_i \mathbf{x} + \sum_{k=1}^{n}\sum_{\ell=1}^{n} G_{ik} G_{j\ell} w_{(k\ell)} \ge 0 \right\},$$

where $w_{(k\ell)} \equiv w_{k\ell}$ if $k \le \ell$ and $w_{(k\ell)} \equiv w_{\ell k}$ otherwise.

This RLT process produces the following linear programming relaxation (as established by Lemma 11.2.5) of Problem NQP, where we have rewritten the objective function of NQP under the linearization (11.16) in a more succinct form, using the symmetry of **H**, and where the notation LP(Ω) emphasizes that the constraints of this problem are (partly) predicated on the hyperrectangle Ω defining NQP. (In the sequel, we shall be partitioning this hyperrectangle.)

$$\text{LP}(\Omega)\text{: Minimize } \sum_{k=1}^{n} c_k x_k + \frac{1}{2}\sum_{k=1}^{n} h_{kk} w_{kk} + \sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n} h_{k\ell} w_{k\ell} \qquad (11.17a)$$

$$\text{subject to } [(g_i - \mathbf{G}_i\mathbf{x})(g_j - \mathbf{G}_j\mathbf{x})]_L \ge 0 \text{ for } 1 \le i \le j \le m. \qquad (11.17b)$$

The following result establishes the relationship between LP(Ω) and its parent problem NPQ. In general, throughout our discussion, for any optimization Problem P, we denote its optimal objective value by $\nu[\text{P}]$.

11.2.5 Lemma

a. Let $\bar{\mathbf{x}}$ be any feasible solution to Problem NQP, and let $\bar{\mathbf{w}}$ be defined according to (11.16) (i.e., $\bar{w}_{k\ell} \equiv \bar{x}_k \bar{x}_\ell$ for $1 \le k \le \ell \le n$). Then $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ is feasible to LP(Ω) and yields the same objective value as in NQP. Hence, in particular, $\nu[\text{LP}(\Omega)] \le \nu[\text{NQP}]$.
b. Conversely, let $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ be any feasible solution to LP(Ω). Then $\bar{\mathbf{x}}$ is feasible to NQP. Moreover, if $(\mathbf{x}^*, \mathbf{w}^*)$ solves LP(Ω) and satisfies the restrictions (11.16), then $\mathbf{x}^*$ solves Problem NQP.

Proof

Part a of the lemma follows directly by the RLT construction process. To establish Part b, suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ is feasible to LP(Ω). Consider any constraint $\mathbf{G}_i\mathbf{x} \le g_i$ defining NQP in (11.15). Note that for any bounding restriction $\ell_j \le x_j \le u_j$ in Ω, the constraint set (11.17b) includes the restrictions

$$[(u_j - x_j)(g_i - \mathbf{G}_i\mathbf{x})]_L \ge 0 \qquad \text{and} \qquad [(x_j - \ell_j)(g_i - \mathbf{G}_i\mathbf{x})]_L \ge 0.$$

Summing these two restrictions, we obtain $(u_j - \ell_j)\ (g_i - \mathbf{G}_i\mathbf{x}) \ge 0$, and so, because $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ is feasible to LP(Ω), we have that $\bar{\mathbf{x}}$ satisfies $g_i - \mathbf{G}_i(\mathbf{x}) \ge 0$. Therefore, $\bar{\mathbf{x}}$ is feasible to NQP. Moreover, if $(\mathbf{x}^*, \mathbf{w}^*)$ solves LP(Ω), and if this solution satisfies (11.16), $\mathbf{x}^*$ is feasible to NQP and yields an objective value equal to $\nu[\text{LP}(\Omega)]$. But $\nu[\text{LP}(\Omega)] \le \nu[\text{NQP}]$ from Part a. Hence, $\mathbf{x}^*$ solves NQP, and this completes the proof.

Lemma 11.2.5 asserts that the constraints (11.17b) imply the original restrictions of Problem NQP, which have therefore been omitted from the representation of Problem LP(Ω). Moreover, the linear program LP(Ω) affords a relaxation of the quadratic program NQP, and if its optimal solution happens to satisfy (11.16), then it also solves the latter problem. The key toward inducing this phenomenon to occur is embodied by the following result.

11.2.6 Lemma

Let $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ be any feasible solution to LP(Ω). Suppose that $\bar{x}_p = \ell_p$ or $\bar{x}_p = u_p$ for some $p \in \{1,..., n\}$. Then $\bar{w}_{(pq)} = \bar{x}_p \bar{x}_q$ for all $q = 1,..., n$.

Proof

Suppose that $\bar{x}_p = \ell_p$ in any feasible solution $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ to LP(Ω), for some $p \in \{1,..., n\}$. Consider any $q \in \{1,..., n\}$. Note that the restrictions (11.17b) include the following constraints:

$$[(x_p - \ell_p)(x_q - \ell_q)]_L \geq 0 \qquad \text{and} \qquad [(x_p - \ell_p)(u_q - x_q)]_L \geq 0.$$

By definition, these inequalities can be rewritten as follows:

$$\ell_q(x_p - \ell_p) + \ell_p x_q \leq w_{(pq)} \leq \ell_p x_q + u_q(x_p - \ell_p).$$

Substituting for $\bar{x}_p$, $\bar{x}_q$, and $\bar{w}_{(pq)}$ above, we get that $\bar{w}_{(pq)} = \ell_p \bar{x}_q = \bar{x}_p \bar{x}_q$. The case for $\bar{x}_p = u_p$ is similar, and this completes the proof.

Observe that Lemma 11.2.6 reveals that for any feasible solution to LP(Ω), if any variable x_p takes on a value at either of its bounds in Ω, then the related new RLT variable $w_{(pq)}$, for each $q = 1,..., n$, will faithfully reproduce the nonlinear product $x_p x_q$ that it represents. This feature is exploited in designing a *branch-and-bound algorithm* for solving Problem NQP. In this procedure, described formally below, we maintain a list of *active nodes* indexed by $q \in Q_s$ at *stage* s of the algorithm, where each node q is associated with some partitioned hyperrectangle $\Omega^q \subseteq \Omega$. To begin with, at Stage $s = 1$, we will have $Q_s \equiv \{1\}$, with $\Omega^1 \equiv \Omega$. Inductively, at any Stage s, given Q_s, we will have computed a lower bound $\text{LB}_q \equiv v[\text{LP}(\Omega^q)]$ (see Lemma 11.2.5) via a construction of Problem (11.17) corresponding to the bounding restrictions imposed by Ω^q. As a result, the lower bound on the original Problem NQP at Stage s is given by $\text{LB}(s) \equiv \min\{\text{LB}_q : q \in Q_s\}$. Furthermore, by Lemma 11.2.5, the solution of each problem of the type LP(Ω^q) produces a feasible solution for NQP. Hence, we can compute its objective value in NQP and thereby retain the best or *incumbent solution* $\mathbf{x}^*$ having an objective value v^*. In case $\text{LB}_q \geq v^*$, we can *fathom* node q (i.e., eliminate it from further consideration) because we know that the corresponding quadratic program defined over $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{x} \in \Omega^q$ cannot yield a solution better than our presently available incumbent solution $\mathbf{x}^*$. Hence, for each Stage s, the active nodes satisfy $\text{LB}_q < v^*$, for all $q \in Q_s$. We now select an active node $q(s) \in Q_s$ that yields the *least lower bound* among the nodes $q \in Q_s$; that is,

$$\text{LB}_{q(s)} = \text{LB}(s) \equiv \min\{\text{LB}_q : q \in Q_s\}.$$

We now proceed to *partition* the corresponding hyperrectangle $\Omega^{q(s)}$ into two *subhyperrectangles*, called its *children hyperrectangles*, based on a *branching variable* x_p selected according to the following rule.

Branching Rule:

Let $(\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$ be the optimal solution obtained for Problem LP($\Omega^{q(s)}$). For ease in notation, denote $(\bar{\mathbf{x}}, \bar{\mathbf{w}}) \equiv (\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$. Compute a *discrepancy index*

$$\theta_k \equiv \max\{0,\ h_{kk}(\bar{x}_k^2 - \bar{w}_{kk})\} + \sum_{\ell=1}^{n} \max\{0,\ h_{k\ell}(\bar{x}_k \bar{x}_\ell - \bar{w}_{k\ell})\} \qquad \text{for } k = 1,\dots,n, \tag{11.18a}$$

and determine the *branching variable* x_p, where the index p corresponds to

$$\theta_p = \max\{\theta_k, k = 1,\dots,n\}. \tag{11.18b}$$

Accordingly, partition $\Omega^{q(s)}$ into two subhyperrectangles by splitting the current bounding interval for x_p within $\Omega^{q(s)}$, say, $\ell_p^{q(s)} \le x_p \le u_p^{q(s)}$, at the value $\bar{x}_p \equiv x_p^{q(s)}$ corresponding to the optimal solution obtained for Problem LP($\Omega^{q(s)}$). This would yield the following two bounding restrictions on x_p, one within each of the resulting children hyperrectangles:

$$\ell_p^{q(s)} \le x_p \le x_p^{q(s)} \equiv \bar{x}_p \quad \text{and} \quad \bar{x}_p \equiv x_p^{q(s)} \le x_p \le u_p^{q(s)}. \tag{11.18c}$$

Note that because $\text{LB}_{q(s)} < v^*$, we must have

$$\theta_p > 0 \quad \text{and} \quad \ell_p^{q(s)} < x_p^{q(s)} < u_p^{q(s)}. \tag{11.18d}$$

This follows because otherwise, if $\theta_p = 0$, then by (11.18a, b) we would have $\theta_k = 0$ for all $k = 1,\dots, n$, so, by (11.18a), this would mean that for each $k = 1,\dots, n$,

$$h_{kk}\bar{x}_k^2 \le h_{kk}\bar{w}_{kk} \qquad \text{and} \qquad h_{k\ell}\bar{x}_k\bar{x}_\ell \le h_{k\ell}\bar{w}_{(k\ell)} \qquad \text{for } \ell = 1,\dots, n, \tag{11.18e}$$

or that the objective value in NQP for the solution $\bar{\mathbf{x}}$ is less than or equal to $v[\text{LP}(\Omega^{q(s)})] = \text{LB}_{q(s)}$, which contradicts $v^* > \text{LB}_{q(s)}$. Moreover, since $\theta_p > 0$ means that at least one of the inequalities in (11.18e) holds true as a reverse strict inequality, so, by Lemma 11.2.6, we must have (11.18d) holding true.

The foregoing analysis motivates the design of the branching rule, which is geared toward identifying the variable that contributes the most to the discrepancy between the new RLT variables that contain it and the associated corresponding nonlinear products that these RLT variables represent. The idea is to drive all such discrepancies to zero. A formal statement of a procedure that accomplishes this is given below.

RLT Algorithm to Solve Problem NQP

Step 0. **Initialization.** Set $s = 1$, $Q_s = \{1\}$, $q(s) = 1$, and $\Omega^1 \equiv \Omega$. Solve $\text{LP}(\Omega^1)$ and let $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ be the solution obtained of objective value $\text{LB}_1 = \nu[\text{LP}(\Omega^1)]$. Initialize the incumbent solution $\mathbf{x}^* = \mathbf{x}^1$, and let the incumbent objective value $v^* = \mathbf{c}^t\mathbf{x}^* + (1/2)\mathbf{x}^{*t}\mathbf{H}\mathbf{x}^*$. If $\text{LB}_1 + \varepsilon \geq v^*$, for some chosen optimality tolerance $\varepsilon \geq 0$, then stop with $\mathbf{x}^*$ as the prescribed solution to Problem NQP. Otherwise, determine a branching variable x_p by using (11.18a, b) and note by (11.18d) that we must have $\theta_p > 0$. Go to Step 1.

Step 1. **Partitioning Step.** Partition the selected active node $\Omega^{q(s)}$ into two subhyperrectangles by splitting the current bounding interval for x_p at the value $\bar{x}_p$ as in (11.18c). Replace $q(s)$ by the node indices for these two new children hyperrectangles to revise Q_s.

Step 2. **Bonnding Step.** Solve the RLT linear programming relaxation for each of the two new nodes generated. Update the incumbent solution if possible. Determine a corresponding branching variable index using (11.18a, b) for each of these two new nodes, as done for Node 1 in the initialization step.

Step 3. **Fathoming Step.** Fathom any nonimproving nodes by setting $Q_{s+1} = Q_s - \{q \in Q_s : \text{LB}_q + \varepsilon \geq v^*\}$. If $Q_{s+1} = \varnothing$, then stop with the prescribed solution $\mathbf{x}^*$ to Problem NQP. Otherwise, increment s by one and go to Step 4.

Step 4. **Node Selection Step.** Select an active node $q(s) \in \text{argmin}\{\text{LB}_q : q \in Q_s\}$, and go to Step 1.

Convergence Analysis for the RLT Algorithm

11.2.7 Theorem

The above RLT algorithm (run with $\varepsilon \equiv 0$) either terminates finitely with the incumbent solution being optimal to Problem NQP, or else an infinite sequence of stages is generated such that along any infinite branch of the branch-and-bound tree, any accumulation point of the x-variable part of the sequence of linear programming relaxation solutions generated for the node subproblems solves NQP.

Proof

The case of finite termination is clear; hence, suppose that an infinite sequence of stages is generated. Consider any infinite branch of the branch-and-bound tree, and suppose that this corresponds to the nested sequence of partitions $\{\Omega^{q(s)}\}$, for stages s belonging to some index set S. For each node $q(s)$, $s \in S$, let $\Omega^{q(s)} \equiv \{\mathbf{x} : \ell^{q(s)} \le \mathbf{x} \le \mathbf{u}^{q(s)}\}$, denote $(\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$ as the optimal solution obtained for LP($\Omega^{q(s)}$), and let $\theta^{q(s)} \equiv (\theta_k^{q(s)},\ k = 1,\dots, n)$ denote the discrepancy index vector determined via (11.18a, b) for the solution $(\bar{\mathbf{x}}, \bar{\mathbf{w}}) \equiv (\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$. By taking any convergent subsequence, if necessary, using the boundedness of the sequences generated, assume without loss of generality that

$$\{(\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)}, \ell^{q(s)}, \mathbf{u}^{q(s)}, \theta^{q(s)})\}_S \to (\hat{\mathbf{x}}, \hat{\mathbf{w}}, \hat{\ell}, \hat{\mathbf{u}}, \hat{\theta}).$$

Note that by the continuity of the constraint functions in (11.17), $(\hat{\mathbf{x}}, \hat{\mathbf{w}})$ is feasible to LP($\hat{\Omega}$), where $\hat{\Omega} \equiv \{\mathbf{x}\colon \hat{\ell} \le \mathbf{x} \le \hat{\mathbf{u}}\}$. Hence, since $\hat{\Omega} \subseteq \Omega$, by Lemma 11.2.5, we have that $\hat{\mathbf{x}}$ is feasible to NQP. We must show that $\hat{\mathbf{x}}$ solves Problem NQP.

Now observe that over the infinite sequence of nodes $\Omega^{q(s)}$, $s \in S$, there exists a variable x_p that is branched on infinitely often via the branching rule (11.18a, b). Let $S_1 \subseteq S$ be the subsequence of stages for which this occurs. By virtue of the partitioning scheme and the nested sequence of hyperrectangles, we know from (11.18c, d) that for each $s \in S_1$, $x_p^{q(s)} \in (\ell_p^{q(s)}, u_p^{q(s)})$, and $x_p^{q(s)} \notin (\ell_p^{q(s')}, u_p^{q(s')})$ for all $s' \in S_1$, $s' > s$. But $[\ell_p^{q(s)}, u_p^{q(s)}] \to [\hat{\ell}_p, \hat{u}_p]$ and $\{x_p^{q(s)}\} \to \bar{x}_p \in [\hat{\ell}_p, \hat{u}_p]$ because $(\hat{\mathbf{x}}, \hat{\mathbf{w}})$ is feasible to LP($\hat{\Omega}$) from above. Consequently, we must have that $\hat{x}_p = \hat{\ell}_p$ or $\hat{x}_p = \hat{u}_p$. By Lemma 11.2.6 and since $(\hat{\mathbf{x}}, \hat{\mathbf{w}})$ is feasible to LP($\hat{\Omega}$), this in turn means that $\hat{w}_{(pq)} = \hat{x}_p \hat{x}_q$, for all $q = 1,\dots, n$. Hence, by (11.18a), we have that $\hat{\theta}_p = 0$. But note that for each $s \in S_1$, we have $\theta_p^{q(s)} \ge \theta_k^{q(s)} \ge 0$ for all $k = 1,\dots, n$, so $0 = \hat{\theta}_p \ge \hat{\theta}_k \ge 0$ for all $k = 1,\dots, n$ (i.e., $\hat{\theta}_k = 0$ for all $k = 1,\dots, n$). As in (11.18e), denoting the objective function of LP($\cdot$) in (11.17a) by $f_{\text{LP}}(\mathbf{x}, \mathbf{w})$, this implies that

$$c^t\hat{x}+\frac{1}{2}\hat{x}^t H\hat{x} \le f_{LP}(\hat{x}, \hat{w}). \tag{11.19a}$$

But by Lemma 11.2.5 and our least lower bound node selection rule, since $f_{LP}(x^{q(s)}, w^{q(s)}) = v[LP(\Omega^{q(s)})] = LB(s) \le v[NQP]$, $\forall\ s \in S$, taking limits as $s \to \infty$, $s \in S$, we get

$$f_{LP}(\hat{x}, \hat{w}) \le v[NQP]. \tag{11.19b}$$

Putting (11.19a) and (11.19b) together, and noting from above that $\hat{x}$ is feasible to NQP, we deduce that

$$v[NQP] \le c^t\hat{x}+\frac{1}{2}x^t H\hat{x} \le f_{LP}(\hat{x}, \hat{w}) \le v[NQP], \tag{11.19c}$$

so equality holds true throughout (11.19c). This means that $\hat{x}$ solves NQP, and this completes the proof.

11.2.8 Example

Consider the following concave minimization quadratic programming problem:

$$\begin{aligned} \text{NQP: Minimize } & -(x_1-12)^2-x_2^2 \\ \text{subject to } & -3x_1+4x_2 \le 24 \\ & 3x_1+8x_2 \le 120 \end{aligned}$$

$$x \in \Omega = \{x : 0 \le x_1 \le 24, 0 \le x_2 \le 15\}.$$

We initialize the RLT algorithm at stage $s = 1$ with $Q_s = \{1\}$ and with $\Omega^1 = \Omega$. The initial relaxation LP(Ω^1) has the objective function (11.17a) to minimize $[-w_{11}-w_{22}+24x_1-144]$. Its constraints (11.17b) are comprised of 10 *bound-factor product inequalities* comprised of pairwise products of the bound factors $(x_j-\ell_j) \ge 0$ and $(u_j-x_j) \ge 0, j = 1, 2$ (including self-products), eight *bound-constraint-factor product inequalities* comprised of multiplying each bound factor with each structural constraint, and three *constraint-factor product inequalities* comprised of pairwise products of the structural constraints (including self-products). (Actually, it can be verified that the bound restriction $x_2 \le 15$ is implied by the other constraints, and hence, it can be omitted from the foregoing RLT constraint generation process because it will simply produce redundant inequalities of the type (11.17b)—see Exercise 11.23.)

For example, one bound-factor product constraint is $[(24-x_1)(x_2)]_L \ge 0$ (i.e., $24x_2 - w_{12} \ge 0$), one bound-constraint-factor product constraint is $[x_1(24+3x_1-4x_2)]_L \ge 0$ (i.e., $24x_1 + 3w_{11} - 4w_{12} \ge 0$), and one constraint-factor product constraint is $[(24 + 3x_1 - 4x_2)^2]_L \geq 0$ (i.e., $9w_{11} + 16w_{22} + 144x_1 - 192x_2 - 24w_{12} + 576 \geq 0$). Solving LP($\Omega^1$), we obtain an optimal solution $(\bar{x}_1, \bar{x}_2, \bar{w}_{11}, \bar{w}_{12}, \bar{w}_{22}) = (8, 6, 192, 48, 72)$ with $v[\text{LP}(\Omega^1)] = -216$. Note that (8, 6) is feasible to NQP (see Lemma 11.6.5) and yields an objective value of -52. Hence, currently, we have $\mathbf{x}^* = (8, 6)$, $v^* = -52$, and $\text{LB}_1 = -216$.

Furthermore, observe that $\bar{w}_{12} = \bar{x}_1\bar{x}_2$, but $\bar{w}_{11} = 192 \neq \bar{x}_1^2 = 64$ and $\bar{w}_{22} = 72 \neq \bar{x}_2^2 = 36$. Hence, we need to partition the current node hyperrectangle by either splitting the interval for x_1 at $\bar{x}_1 = 8$ or that for x_2 at $\bar{x}_2 = 6$. To make this choice we resort to the **branching** rule (11.18a, b). Using (11.18a), we first compute $\theta_1 = \max\{0, -(64 - 192)\} = 128$ and $\theta_2 = \max\{0, -(36 - 72)\} = 36$. Hence, from (11.18b), we select $x_p \equiv x_1$, and by (11.18c), we create two children hyperrectangles to replace Ω^1, as given by

$$\Omega^2 = \{\mathbf{x} : 0 \leq x_1 \leq 8, 0 \leq x_2 \leq 15\}$$

$$\text{and} \qquad \Omega^3 = \{\mathbf{x} : 8 \leq x_1 \leq 24, 0 \leq x_2 \leq 15\}.$$

Tentatively, we accordingly revise $Q_1 = \{2, 3\}$ at Step 1. The reader can now verify (see Exercise 11.24) that at Step 2 we obtain $v[\text{LP}(\Omega^2)] = v[\text{LP}(\Omega^3)] = -180$, with the x-part of the respective LP solutions being $(0, 6)^t$ and $(24, 6)^t$, both yielding an objective value of -180 in Problem NQP. Hence, the algorithm can be terminated since we will therefore obtain $Q_2 = \varnothing$ at Step 3.

In concluding this section we direct the reader's attention to Exercise 11.25, where it is shown that by constructing selected quadratic as well as *cubic RLT constraints*, we can construct an LP relaxation for this example that directly solves the given quadratic program at the initial node itself, without requiring further branching. In general, various such enhancements have been proposed to accelerate the convergence of the algorithm based on generating an appropriate filtered set of valid inequalities (including convex nonlinear restrictions), applying *semidefinite programming* concepts (including the generation of related *semidefinite cuts*) (see Exercise 11.26), implementing alternative branching strategies, tightening bound restrictions in a preprocessing step via feasibility plus optimality considerations, and scaling along with possibly using affine transformations on the original problem to improve its structural properties. These mechanisms can also be applied to solve wider classes of polynomial (see Exercise 11.27), factorable, and black-box optimization problems. We refer the reader to the Notes and References section for a more detailed study of this subject.

11.3 Separable Programming

In this section we discuss the use of the simplex method to obtain solutions to nonlinear programs where the objective function and the constraint functions can be expressed as the sum of functions, each involving only one variable. We denote such a *separable nonlinear program* as Problem P and express it as follows:

$$\begin{aligned} \text{P:}\quad & \text{Minimize} \quad \sum_{j=1}^{n} f_j(x_j) \\ & \text{subject to} \quad \sum_{j=1}^{n} g_{ij}(x_j) \le p_i \qquad \text{for } i = 1,\ldots,m \\ & \qquad\qquad\qquad x_j \ge 0 \qquad \text{for } j = 1,\ldots,n. \end{aligned} \tag{11.20}$$

Problems of this type arise in numerous applications, including econometric data fitting, electrical network analysis, design and management of water supply systems, logistics, and statistics.

Approximating the Separable Problem

We now discuss how we can define a new problem that approximates the original Problem P. The new problem is obtained by replacing each nonlinear function by an approximating piecewise linear function. To see how this can be done, consider a continuous function θ of the variable μ. Suppose that we are interested in values of θ over the interval $[a, b]$. We wish to define a piecewise linear function $\hat{\theta}$ that approximates θ. The interval $[a, b]$ is first partitioned into smaller intervals, via the grid points $a = \mu_1, \mu_2, \ldots, \mu_k = b$, as shown in Figure 11.3. The function θ is approximated in the interval $[\mu_v, \mu_{v+1}]$ as follows. Let $\mu = \lambda\mu_v + (1 - \lambda)\mu_{v+1}$ for some $\lambda \in [0, 1]$. Then

$$\hat{\theta}(\mu) = \lambda\theta(\mu_v) + (1 - \lambda)\theta(\mu_{v+1}). \tag{11.21}$$

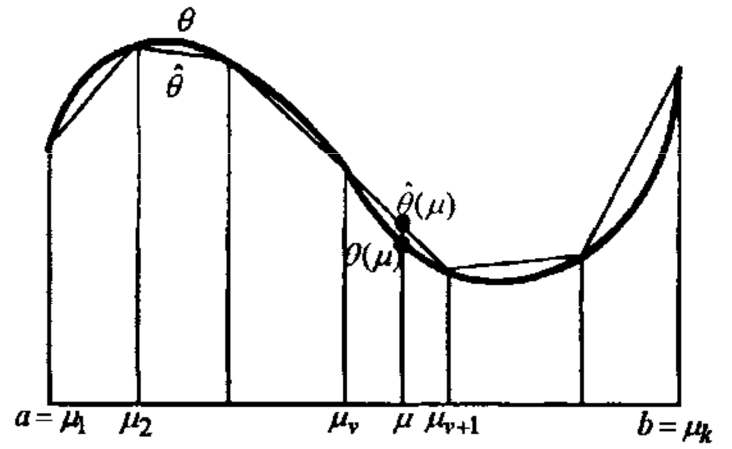
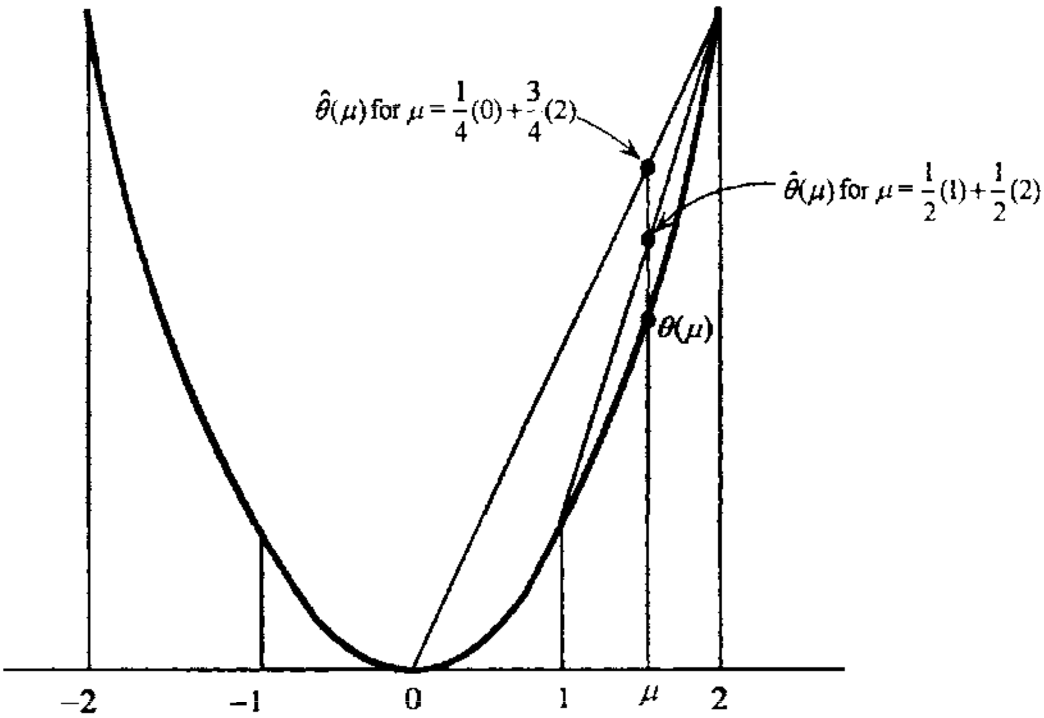


Figure 11.3 Piecewise linear approximation of a function.

Note that the grid points may or may not be equidistant, and that the accuracy of the approximation improves as the number of grid points increases. Note, however, that a major difficulty may arise in using the foregoing linear approximation to a function. This is because a given point μ in the interval $[\mu_v, \mu_{v+1}]$ can alternatively be represented as a convex combination of two or more *nonadjacent* grid points. To illustrate, consider the function θ defined by $\theta(\mu) = \mu^2$. The graph of the function on the interval [–2, 2] is shown in Figure 11.4. Suppose that we use the grid points –2, –1, 0, 1, and 2. The point $\mu = 1.5$ can be written as (1/2)(1) + (1/2)(2) and also as (1/4)(0) + (3/4)(2). The value of the function θ at $\mu = 1.5$ is 2.25. The first approximation gives $\hat{\theta}(\mu) = (1/2)\theta(1) + (1/2)\theta(2) = 2.5$, whereas the second approximation gives $\hat{\theta}(\mu) = (1/4)\theta(0) + (3/4)\theta(2) = 3$. Clearly, the first approximation using adjacent grid points yields a better approximation. In general, therefore, the function θ can be approximated over the interval $[a, b]$ via the grid points $\mu_1, \ldots, \mu_k$ by the piecewise linear function $\hat{\theta}$, defined by

$$\hat{\theta}(\mu) = \sum_{v=1}^{k} \lambda_v \theta(\mu_v), \ \sum_{v=1}^{k} \lambda_v = 1, \ \lambda_v \geq 0 \qquad \text{for } v = 1, \ldots, k \tag{11.22}$$

where at most two λ_v-variables are positive, and they must be adjacent. This representation is known as the *λ-form approximation*. An alternative related representation, known as the *δ-form approximation*, is described in Exercises 11.35 and 11.36.



Fignre 11.4 Importance of adjacency in approximation.

We now present a problem that approximates the separable Problem P defined by (11.19). This is done by considering each variable x_j for which either f_j or g_{ij} is nonlinear for some $i = 1,\ldots, m$ and replacing it with the piecewise linear approximation defined by (11.22). For the sake of clarity, we define a set L as

$$L = \{j : f_j \text{ and } g_{ij} \text{ for } i = 1,\ldots, m \text{ are linear}\}.$$

Then, for each $j \notin L$, we consider the interval of interest $[a_j, b_j]$, where $a_j, b_j \geq 0$. We can now define the grid points x_{vj} for $v = 1,\ldots, k_j$, where $x_{1j} = a_j$ and $x_{k_j j} = b_j$. Note that the grid points need not be spaced equally and that different grid lengths could be used for different variables. However, from Theorem 11.3.4, as will be seen later, the maximum grid length used is related to the accuracy of the solution obtained. Using the grid points for each $j \notin L$, from (11.22), the functions f_j and g_{ij} for $i = 1,\ldots, m$ could be replaced by their linear approximations

$$\hat{f}_j(x_j) = \sum_{v=1}^{k_j} \lambda_{vj} f(x_{vj}) \qquad \text{for } j \notin L$$

$$\hat{g}_{ij}(x_j) = \sum_{v=1}^{k_j} \lambda_{vj} g_{ij}(x_{vj}) \qquad \text{for } i = 1,\ldots, m \text{ and } j \notin L$$

$$\sum_{v=1}^{k_j} \lambda_{vj} = 1 \qquad \text{for } j \notin L$$

$$\lambda_{vj} \geq 0 \qquad \text{for } v = 1,\ldots, k_j \text{ and } j \notin L.$$

By definition, both f_j and g_{ij} for $i = 1,\ldots, m$ are linear for $j \in L$. Hence, no grid points need be defined, and in this case, the linear approximations are given by

$$\hat{f}_j(x_j) \equiv f_j(x_j),\ \hat{g}_{ij}(x_j) \equiv g_{ij}(x_j) \quad \text{for } i = 1,\ldots,m \text{ and } j \in L.$$

The following Problem AP can then be viewed as the problem that approximates the original Problem P.

$$\begin{aligned} \text{AP: Minimize} \quad & \sum_{j\in L} f_j(x_j) + \sum_{j\notin L} \hat{f}_j(x_j) \\ \text{subject to} \quad & \sum_{j\in L} g_{ij}(x_j) + \sum_{j\notin L} \hat{g}_{ij}(x_j) \leq p_i \quad \text{for } i = 1,\ldots, m \qquad (11.23) \\ & x_j \geq 0 \quad \text{for } j = 1,\ldots, n. \end{aligned}$$

Note that the objective function and constraints in Problem AP are piecewise linear. However, by using the definitions of $\hat{f}_j$ and $\hat{g}_{ij}$ for $j \notin L$, the problem can be restated in an equivalent more manageable form as Problem LAP:

$$\begin{aligned}
\text{LAP: Minimize } & \sum_{j\in L} f_j(x_j) + \sum_{j\notin L} \sum_{v=1}^{k_j} \lambda_{vj} f_j(x_{vj}) \\
\text{subject to } & \sum_{j\in L} g_{ij}(x_j) + \sum_{j\notin L} \sum_{v=1}^{k_j} \lambda_{vj} g_{ij}(x_{vj}) \le p_i \quad \text{for } i = 1,\dots,m \\
& \sum_{v=1}^{k_j} \lambda_{vj} = 1 \quad \text{for } j \notin L \\
& \lambda_{vj} \ge 0 \quad \text{for } v = 1,\dots,k_j;\ j \notin L \\
& x_j \ge 0 \quad \text{for } j \in L
\end{aligned} \tag{11.24}$$

At most, two adjacent λ_{vj}-values are positive for $j \notin L$.

Solving the Approximating Problem

With the exception of the constraint that, at most, two adjacent λ_{vj} variables are positive for $j \notin L$, Problem LAP is a linear program. For solving Problem LAP, we can use the simplex method with the following *restricted basis entry rule.* A nonbasic variable λ_{vj} is introduced into the basis only if it improves the objective function and if the new basis has no more than two adjacent λ_{vj} variables that are positive for each $j \notin L$. Theorem 11.3.1 shows that for $j \notin L$, if g_{ij} is convex for $i = 1,\dots, m$ and if f_j is strictly convex, we can discard the restricted basis entry rule and adopt the simplex method for linear programming as described in Section 2.7.

11.3.1 Theorem

Consider Problem P to minimize $\sum_{j=1}^{n} f_j(x_j)$ subject to $\sum_{j=1}^{n} g_{ij}(x_j) \le p_i$ for $i = 1,\dots, m$, and $x_j \ge 0$ for $j = 1,\dots, n$. Let $L = \{j : f_j$ and g_{ij} for $i = 1,\dots, m$ are linear$\}$. Assume that for $j \notin L$, f_j is strictly convex and that g_{ij} is convex for $i = 1,\dots, m$. Suppose further that for each $j \notin L$, f_j and g_{ij} for $i = 1,\dots, m$ are replaced by their piecewise linear approximations via the grid points x_{vj} for $v = 1,\dots, k_j$, yielding the linear program defined below.

$$\begin{aligned}
\text{Minimize} \quad & \sum_{j\in L} f_j(x_j)+\sum_{j\notin L}\sum_{v=1}^{k_j}\lambda_{vj}f_j(x_{vj}) \\
\text{subject to} \quad & \sum_{j\in L} g_{ij}(x_j)+\sum_{j\notin L}\sum_{v=1}^{k_j}\lambda_{vj}g_{ij}(x_{vj})\le p_i \quad \text{for } i=1,\dots,m \\
& \sum_{v=1}^{k_j}\lambda_{vj}=1 \quad \text{for } j\notin L \\
& \lambda_{vj}\ge 0 \quad \text{for } v=1,\dots,k_j;\ j\notin L \\
& x_j\ge 0 \quad \text{for } j\in L.
\end{aligned} \tag{11.25}$$

Let $\hat{x}_j$ for $j \in L$ and $\hat{\lambda}_{vj}$ for $v = 1,\dots,k_j$ and $j \notin L$ solve the above problem. Then:

1. For each $j \notin L$, at most two $\hat{\lambda}_{vj}$-values are positive, and they are necessarily adjacent.
2. Let $\hat{x}_j = \sum_{v=1}^{k_j}\hat{\lambda}_{vj}x_{vj}$ for $j \notin L$. Then the vector $\hat{\mathbf{x}}$ whose jth component is $\hat{x}_j$ for $j = 1,\dots, n$ is feasible to Problem P.

Proof

To prove Part 1 it suffices to show that for each $j \notin L$, if $\hat{\lambda}_{\ell j}$ and $\hat{\lambda}_{pj}$ are positive, the grid points $x_{\ell j}$ and x_{pj} are necessarily adjacent. By contradiction, suppose that there exist $\hat{\lambda}_{\ell j}$ and $\hat{\lambda}_{pj} > 0$, where $x_{\ell j}$ and x_{pj} are not adjacent. Then there exists a grid point $x_{\gamma j} \in (x_{\ell j}, x_{pj})$ that can be expressed as $x_{\gamma j} = \alpha_1 x_{\ell j} + \alpha_2 x_{pj}$, where $\alpha_1, \alpha_2 > 0$ and $\alpha_1 + \alpha_2 = 1$. Now, consider the optimal solution to the problem defined by (11.25). Let $u_i \ge 0$ for $i = 1,\dots, m$ be the optimum Lagrangian multipliers associated with the first m constraints, and for each $j \notin L$, let v_j be the optimal Lagrangian multiplier associated with the constraint $\sum_{v=1}^{k_j}\lambda_{vj} = 1$. Then the following subset of the KKT necessary conditions are satisfied:

$$f_j(x_{\ell j})+\sum_{i=1}^{m}u_i g_{ij}(x_{\ell j})+v_j=0 \tag{11.26}$$

$$f_j(x_{pj})+\sum_{i=1}^{m}u_i g_{ij}(x_{pj})+v_j=0 \tag{11.27}$$

$$f_j(x_{vj})+\sum_{i=1}^{m}u_i g_{ij}(x_{vj})+v_j\ge 0 \qquad \text{for } v=1,\dots,k_j. \tag{11.28}$$

We show below that the last condition is contradicted for $\nu = \gamma$. By the strict convexity of f_j, the convexity of g_{ij}, and by (11.26) and (11.27), we have

$$f_j(x_{\gamma j}) + \sum_{i=1}^{m} u_i g_{ij}(x_{\gamma i}) + v_j < \alpha_1 f_j(x_{\ell j}) + \alpha_2 f_j(x_{pj})$$
$$+ \sum_{i=1}^{m} u_i[\alpha_1 g_{ij}(x_{\ell j}) + \alpha_2 g_{ij}(x_{pj})] + v_j = 0.$$

This contradicts (11.28) for $\nu = \gamma$, and hence, $x_{\ell j}$ and x_{pj} must be adjacent, and Part 1 of the theorem is proved.

To prove Part 2, from the convexity of g_{ij} for $j \notin L$ and for each $i = 1,\ldots, m$, and noting that $\hat{x}_j$ for $j \in L$, and $\hat{\lambda}_{\nu j}$ for $\nu = 1,\ldots, k_j$, $j \notin L$, satisfy the constraints in (11.25), we get

$$\begin{aligned} g_i(\hat{\mathbf{x}}) &= \sum_{j\in L} g_{ij}(\hat{x}_j) + \sum_{j\notin L} g_{ij}(\hat{x}_j) \\ &= \sum_{j\in L} g_{ij}(\hat{x}_j) + \sum_{j\notin L} g_{ij}\left(\sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} x_{\nu j}\right) \\ &\le \sum_{j\in L} g_{ij}(\hat{x}_j) + \sum_{j\notin L} \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} g_{ij}(x_{\nu j}) \\ &\le p_i \end{aligned}$$

for $i = 1,\ldots, m$. Furthermore, $\hat{x}_j \ge 0$ for $j \in L$, and $\hat{x}_j = \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} x_{\nu j} \ge 0$ for $j \notin L$, since $\hat{\lambda}_{\nu j}$, $x_{\nu j} \ge 0$ for $\nu = 1,\ldots, k_j$ and $j \notin L$. Hence, $\hat{\mathbf{x}}$ is feasible to Problem P and the proof is complete.

11.3.2 Example

Consider the following separable program:

$$\begin{aligned} \text{Minimize} \quad & x_1^2 - 6x_1 + x_2^2 - 8x_2 - \frac{1}{2}x_3 \\ \text{subject to} \quad & x_1 + x_2 + x_3 \le 5 \\ & x_1^2 - x_2 \le 3 \\ & x_1, x_2, x_3 \ge 0. \end{aligned}$$

Note that $L = \{3\}$, since there are no nonlinear terms involving x_3, and hence, we will not construct any grid points for x_3. From the constraints it is clear that both x_1 and x_2 must lie in the interval [0, 5]. Recall that the grid points need not be equally spaced. For the variables x_1 and x_2, we use the grid points 0, 2, 4, and

5, so that $x_{11} = 0$, $x_{21} = 2$, $x_{31} = 4$, and $x_{41} = 5$, and $x_{12} = 0$, $x_{22} = 2$, $x_{32} = 4$, and $x_{42} = 5$. Thus,

$$0\lambda_{11} + 2\lambda_{21} + 4\lambda_{31} + 5\lambda_{41} = x_1$$

$$0\lambda_{12} + 2\lambda_{22} + 4\lambda_{32} + 5\lambda_{42} = x_2$$

$$\lambda_{11} + \lambda_{21} + \lambda_{31} + \lambda_{41} = 1$$

$$\lambda_{12} + \lambda_{22} + \lambda_{32} + \lambda_{42} = 1$$

$$\lambda_{v1}, \lambda_{v2} \geq 0 \qquad \text{for } v = 1, 2, 3, 4$$

$$\hat{f}(x) = (-8\lambda_{21} - 8\lambda_{31} + 5\lambda_{41}) + (-12\lambda_{22} - 16\lambda_{32} - 15\lambda_{42}) - \frac{1}{2}x_3$$

$$\hat{g}_1(x) = (2\lambda_{21} + 4\lambda_{31} + 5\lambda_{41}) + (2\lambda_{22} + 4\lambda_{32} + 5\lambda_{42}) + x_3 \leq 5$$

$$\hat{g}_2(x) = (4\lambda_{21} + 16\lambda_{31} + 25\lambda_{41}) - (2\lambda_{22} + 4\lambda_{32} + 5\lambda_{42}) \leq 3.$$

Introducing the slack variables x_4 and x_5, we get the first tableau given below. We solve the problem using the simplex method with the restricted basis entry rule. The sequence of tableaux obtained are given as follows:

	z	λ_{11}	λ_{21}	λ_{31}	λ_{41}	λ_{12}	λ_{22}	λ_{32}	λ_{42}	x_3	x_4	x_5	RHS
z	1	0	8	8	5	0	12	16	15	1/2	0	0	0
x_4	0	0	2	4	5	0	2	4	5	1	1	0	5
x_5	0	0	4	16	25	0	−2	−4	−5	0	0	1	3
λ_{11}	0	1	1	1	1	0	0	0	0	0	0	0	1
λ_{12}	0	0	0	0	0	1	1	①	1	0	0	0	1

	z	λ_{11}	λ_{21}	λ_{31}	λ_{41}	λ_{12}	λ_{22}	λ_{32}	λ_{42}	x_3	x_4	x_5	RHS
z	1	0	8	8	5	−16	−4	0	−1	1/2	0	0	−16
x_4	0	0	②	4	5	−4	−2	0	1	1	1	0	1
x_5	0	0	4	16	25	4	2	0	−1	0	0	1	7
λ_{11}	0	1	1	1	1	0	0	0	0	0	0	0	1
λ_{32}	0	0	0	0	0	1	1	1	1	0	0	0	1

	z	λ_{11}	λ_{21}	λ_{31}	λ_{41}	λ_{12}	λ_{22}	λ_{32}	λ_{42}	x_3	x_4	x_5	RHS
z	1	0	0	−8	−15	0	4	0	−5	−7/2	−4	0	−20
λ_{21}	0	0	1	2	5/2	−2	−1	0	1/2	1/2	1/2	0	1/2
x_5	0	0	0	8	15	12	6	0	−3	−2	−2	1	5
λ_{11}	0	1	0	−1	−3/2	2	①	0	−1/2	−1/2	−1/2	0	1/2
λ_{32}	0	0	0	0	0	1	1	1	1	1	0	0	1

	z	λ_{11}	λ_{21}	λ_{31}	λ_{41}	λ_{12}	λ_{22}	λ_{32}	λ_{42}	x_3	x_4	x_5	RHS
z	1	−4	0	−4	−9	−8	0	0	−3	−3/2	−2	0	−22
λ_{21}	0	1	1	1	1	0	0	0	0	0	0	0	1
x_5	0	−6	0	14	24	0	0	0	0	1	1	1	2
λ_{22}	0	1	0	−1	−3/2	2	1	0	−1/2	−1/2	−1/2	0	1/2
λ_{32}	0	−1	0	1	3/2	−1	0	1	3/2	1/2	1/2	0	1/2

Note that at the second tableau, λ_{31} could not be introduced into the basis, as it would have violated the restricted basis entry rule. From the final tableau, the optimal solution to the approximating Problem AP is $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2\ \hat{x}_3)^t$, where

$$\begin{aligned}\hat{x}_1 &= 2\hat{\lambda}_{21} + 4\hat{\lambda}_{31} + 5\hat{\lambda}_{41} = 2\\ \hat{x}_2 &= 2\hat{\lambda}_{22} + 4\hat{\lambda}_{32} + 5\hat{\lambda}_{42} = 3\\ \hat{x}_3 &= 0.\end{aligned}$$

The corresponding value of the objective function for Problem AP is $\hat{f}(2, 3, 0) = -22$, whereas the value of the objective function for the original Problem P at this point is $f(2, 3, 0) = -23$. Note that the objective function and the constraint functions for this problem satisfy the assumptions of Theorem 11.3.1. Thus, we could have adopted the simplex method without the restricted basis entry rule and yet obtained the above optimal solution.

Relationship Between the Optimal Solutions to the Original and Approximating Problems

As we have seen from Theorem 11.3.1, in the presence of suitable convexity assumptions, an optimal solution to the approximating linear programming problem is a feasible solution to the original problem. We show in Theorem 11.3.4 that if the grid length is chosen sufficiently small, the optimal objective values to both problems could be made arbitrarily close. To prove this result, the following theorem is needed.

11.3.3 Theorem

Consider Problems P and AP defined in (11.20) and (11.23), respectively. For $j \notin L$, suppose that f_j and g_{ij} for $i = 1,\ldots, m$ are convex; and furthermore, let $\hat{f}_j$ and $\hat{g}_{ij}$ represent their piecewise linear approximations on the interval $[a_j, b_j]$. For $j \notin L$ and for $i = 1,\ldots, m$, let c_{ij} be such that $|g'_{ij}(x_j)| \le c_{ij}$ for $x_j \in [a_j, b_j]$. Furthermore, for $j \notin L$, let c_j be such that $|f'_j(x_j)| \le c_j$ for $x_j \in [a_j, b_j]$. For $j \notin L$, let δ_j be the maximum grid length used for the variable x_j. Then

$$\hat{f}(\mathbf{x}) \geq f(\mathbf{x}) \geq \hat{f}(\mathbf{x}) - c$$

$$\hat{g}_i(\mathbf{x}) \geq g_i(\mathbf{x}) \geq \hat{g}_i(\mathbf{x}) - c \qquad \text{for } i = 1,\ldots, m$$

where $c = \max_{0 \leq i \leq m} \{\overline{c}_i\}$ and where

$$\overline{c}_0 = \sum_{j \notin L} 2c_j\delta_j \text{ and } \overline{c}_i = \sum_{j \notin L} 2c_{ij}\delta_j \qquad \text{for } i = 1,\ldots, m.$$

Proof

We first show that $\hat{f}_j(x_j) \geq f_j(x_j) \geq \hat{f}_j(x_j) - 2c_j\delta_j$ for $j \notin L$. Let $j \notin L$, and let $x_j \in [a_j, b_j]$. Then there exist grid points μ_k and μ_{k+1} such that $x_j \in [\mu_k, \mu_{k+1}]$. Furthermore, $x_j = \lambda\mu_k + (1-\lambda)\mu_{k+1}$ for some $\lambda \in [0, 1]$. By the definition of $\hat{f}_j$, and noting the convexity of f_j and that $\lambda \in [0, 1]$, we get

$$\hat{f}_j(x_j) = \lambda f_j(\mu_k) + (1-\lambda) f_j(\mu_{k+1}) \geq f_j(\lambda\mu_k + (1-\lambda)\mu_{k+1}) = f_j(x_j).$$

Now we show that $f_j(x_j) \geq \hat{f}_j(x_j) - 2c_j\delta_j$. Note that $\hat{f}_j(x_j)$ can be represented as follows:

$$\hat{f}_j(x_j) = f_j(\mu_k) + (x_j - \mu_k)s, \tag{11.29}$$

where $s = [f_j(\mu_{k+1}) - f_j(\mu_k)]/[\mu_{k+1} - \mu_k]$. Furthermore, by Theorem 3.3.3, it follows that

$$f_j(x_j) \geq f_j(\mu_k) + (x_j - \mu_k) f_j'(\mu_k). \tag{11.30}$$

Subtracting (11.30) from (11.29), we get

$$\hat{f}_j(x_j) - f_j(x_j) \leq (x_j - \mu_k)[s - f_j'(\mu_k)]. \tag{11.31}$$

By the mean value theorem, there exists a $y \in [\mu_k, \mu_{k+1}]$ such that $s = f_j'(y)$. Thus, by assumption, $s - f_j'(\mu_k) \leq 2c_j$. Furthermore, $x_j - \mu_k \leq \delta_j$, and hence, from (11.31), we must have $\hat{f}_j(x_j) - f_j(x_j) \leq 2c_j\delta_j$. We have thus proved that

$$\hat{f}_j(x_j) \geq f_j(x_j) \geq \hat{f}_j(x_j) - 2c_j\delta_j \qquad \text{for } j \notin L \tag{11.32}$$

for each $x_j \in [a_j, b_j]$. Summing (11.32) over $j \notin L$ and adding $\sum_{j \in L} f_j(x_j)$ to each term, it follows that

$$\hat{f}(\mathbf{x}) \geq f(\mathbf{x}) \geq \hat{f}(\mathbf{x}) - \bar{c}_0. \tag{11.33}$$

In a similar fashion, we get

$$\hat{g}_i(\mathbf{x}) \geq g_i(\mathbf{x}) \geq \hat{g}_i(\mathbf{x}) - \bar{c}_i \qquad \text{for } i = 1,\dots, m. \tag{11.34}$$

By the definition of c, and from (11.33) and (11.34), the result follows.

11.3.4 Theorem

Consider Problem P, defined in (11.20). Let $L = \{j : f_j$ and g_{ij} for $i = 1,\dots, m$ are linear$\}$. For $j \notin L$, let $\hat{f}_j$ and $\hat{g}_{ij}$ be the piecewise linear approximations of f_j and g_{ij}, respectively, for $i = 1,\dots, m$. Let Problem AP, defined in (11.23), and Problem LAP, defined in (11.24), be the equivalent problems that approximate Problem P. For $j \notin L$, suppose that f_j and g_{ij} for $i = 1,\dots, m$ are convex. Let $\bar{\mathbf{x}}$ be an optimal solution to Problem P. Let $\hat{x}_j$ for $j \in L$, and $\hat{\lambda}_{vj}$ for $v = 1,\dots, k_j$, $j \notin L$, be an optimal solution to Problem LAP such that the vector $\hat{\mathbf{x}}$, whose components are $\hat{x}_j$ for $j \in L$ and $\hat{x}_j = \sum_{v=1}^{k_j} \hat{\lambda}_{vj} x_{vj}$ for $j \notin L$, is an optimal solution to Problem AP. Let $\hat{u}_i \geq 0$ be the corresponding optimal Lagrangian multiplier obtained associated with the constraint $\hat{g}_i(\mathbf{x}) \leq p_i$ for $i = 1,\dots, m$. Then:

1. $\hat{\mathbf{x}}$ is a feasible solution to Problem P.
2. $0 \leq f(\hat{\mathbf{x}}) - f(\bar{\mathbf{x}}) \leq c(1 + \sum_{i=1}^{m} \hat{u}_i)$, where c is as defined in Theorem 11.3.3.

Proof

The vector $\hat{\mathbf{x}}$ is feasible to Problem AP; that is, $\hat{g}_i(\hat{\mathbf{x}}) \leq p_i$ for $i = 1,\dots, m$, and $\hat{\mathbf{x}} \geq \mathbf{0}$. By Theorem 11.3.3, $\hat{g}_i(\hat{\mathbf{x}}) \leq p_i$ implies that $g_i(\hat{\mathbf{x}}) \leq p_i$ for $i = 1,\dots, m$, and Part 1 follows.

The reader can verify that a piecewise linear approximation of a convex function is also convex, so that $\hat{f}_j$ and $\hat{g}_{ij}$ are convex for $i = 1,\dots, m$ and $j \notin L$. Since the sum of convex functions is also convex, the objective function and constraint functions of Problem AP are convex. Hence, $(\hat{\mathbf{x}}, \hat{\mathbf{u}})$ satisfies the saddle point optimality criteria of Problem AP, given in Theorem 6.2.5, so that

$$\hat{f}(\hat{\mathbf{x}}) \leq \hat{f}(\mathbf{x}) + \hat{\mathbf{u}}^t[\hat{\mathbf{g}}(\mathbf{x}) - \mathbf{p}] \qquad \text{for all } \mathbf{x} \geq \mathbf{0}. \tag{11.35}$$

Since $g_i(\bar{\mathbf{x}}) \leq p_i$, by Theorem 11.3.3, $\hat{g}_i(\bar{\mathbf{x}}) - p_i \leq c$ for $i = 1,\dots, m$. Letting $\mathbf{x} = \bar{\mathbf{x}}$ in (11.35) and noting that $\hat{\mathbf{u}} \geq \mathbf{0}$, it follows that

$$\hat{f}(\hat{\mathbf{x}}) \le \hat{f}(\bar{\mathbf{x}}) + c\sum_{i=1}^{m} \hat{u}_i. \tag{11.36}$$

By Part 1 of the theorem, $\hat{\mathbf{x}}$ is feasible to Problem P, and hence, $f(\hat{\mathbf{x}}) \ge f(\bar{\mathbf{x}})$. From Theorem 11.3.3, $f(\bar{\mathbf{x}}) \ge \hat{f}(\bar{\mathbf{x}}) - c$, and hence, $f(\hat{\mathbf{x}}) \ge f(\bar{\mathbf{x}}) \ge \hat{f}(\bar{\mathbf{x}}) - c$. From (11.36) and since $\hat{f}(\hat{\mathbf{x}}) \ge f(\hat{\mathbf{x}})$, it follows that

$$f(\hat{\mathbf{x}}) \ge f(\bar{\mathbf{x}}) \ge \hat{f}(\hat{\mathbf{x}}) - c\left(1 + \sum_{i=1}^{m} \hat{u}_i\right) \ge f(\hat{\mathbf{x}}) - c\left(1 + \sum_{i=1}^{m} \hat{u}_i\right).$$

This completes the proof.

In Theorem 11.3.4, the Lagrangian multipliers $\hat{u}_i$ for $i = 1,\ldots, m$ are immediately available from the optimal simplex tableau for Problem LAP. When the approximating problem is solved, we can use Theorem 11.3.4 to determine the maximum deviation $c(1+\sum_{i=1}^{m} \hat{u}_i)$ of the true optimal objective value from that at hand. Note that as the grid length is reduced, c will be smaller, and hence, a better approximation will be obtained. The Notes and References section points the reader to literature on *error estimations* in convex separable programming.

Generation of the Grid Points

It may be noted that the accuracy of the procedure discussed above depends largely on the number of grid points for each variable. However, as the number of grid points is increased, the number of variables in the approximating linear program LAP also increases. One approach is to use a coarse grid initially and then to use a finer grid around the optimal solution obtained with the coarse grid. An attractive alternative is to generate grid points when necessary. This approach is discussed below. (See Meyer [1979, 1980] for an alternative that employs a sequence of two segment approximations only.)

Consider Problem LAP, defined in (11.24). Let x_{vj} for $v = 1,\ldots, k_j, j \notin L$ be the grid points considered so far. Let $\hat{x}_j$ for $j \in L$ and $\hat{\lambda}_{vj}$ for $v = 1,\ldots, k_j, j \notin L$, solve Problem LAP. Furthermore, let $\hat{u}_i \ge 0$ for $i = 1,\ldots, m$ be the optimal Lagrangian multipliers associated with the first m constraints, and let $\hat{v}_j$ for each $j \notin L$ be the Lagrangian multiplier associated with the constraint $\sum_{v=1}^{k_j} \lambda_{vj} = 1$. Note that the solution values $\hat{x}_j$, $\hat{\lambda}_{vj}$, $\hat{u}_i$, and $\hat{v}_j$ satisfy the KKT conditions for Problem LAP. We wish to know whether we need to consider an additional grid point for any of the variables x_j for $j \notin L$ to yield a better piecewise linear approximation in the sense that if this new grid point were considered in defining Problem LAP, its minimum objective function value would decrease. For some $j \notin L$, suppose we were to consider a grid point $x_{\gamma j}$. The reader may verify that if

$$f_j(x_{\gamma j}) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_{\gamma j}) + \hat{v}_j \geq 0, \tag{11.37}$$

then letting $\hat{\lambda}_{\gamma j} = 0$ will satisfy all the KKT conditions for the revised Problem LAP. However, since we do not know where the new grid point is to be located, we can answer the question whether all x_j satisfying $a_j \leq x_j \leq b_j$ for $j \notin L$ will satisfy (11.37) by solving subproblem PS for each $j \notin L$:

$$\begin{aligned} \text{PS:}\quad & \text{Minimize} \quad f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_j) + \hat{v}_j \\ & \text{subject to} \quad a_j \leq x_j \leq b_j. \end{aligned}$$

If the minimum objective function value is nonnegative for all $j \notin L$, then we cannot find a new grid point contradicting (11.37). Theorem 11.3.5 asserts that if this is the case, the current solution is optimal to the original Problem P and that if the minimum objective value is negative for some $j \notin L$, then we can get a better approximation to Problem P. Furthermore, the theorem provides bounds on the optimum objective function value for Problem P at each iteration.

11.3.5 Theorem

Consider Problem P defined in (11.20). Let $L = \{j : f_j$ and g_{ij} for $i = 1,\ldots, m$ are linear$\}$. Suppose, without loss of generality, that $f_j(x_j)$ is of the form $c_j x_j$ and $g_{ij}(x_j)$ is of the form $a_{ij}x_j$ for $i = 1,\ldots, m$ and for $j \in L$. Using the grid points $x_{\nu j}$, $\nu = 1,\ldots, k_j$ for $j \notin L$, let Problem LAP be defined as in (11.24). For $j \notin L$, suppose that f_j and g_{ij} are convex for $i = 1,\ldots, m$. Let $\hat{x}_j$ for $j \in L$, and $\hat{\lambda}_{\nu j}$ for $\nu = 1,\ldots, k_j$, $j \notin L$, be optimal to Problem LAP with a corresponding objective function value $\hat{z}$. Let $\hat{u}_i \geq 0$ for $i = 1,\ldots, m$, be the Lagrangian multipliers corresponding to the first m constraints, and let $\hat{v}_j$ for $j \notin L$ be the Lagrangian multipliers associated with the constraints $\sum_{\nu=1}^{k_j} \lambda_{\nu j} = 1$ in Problem LAP. Now, for each $j \notin L$, consider the following problem:

$$\begin{aligned} & \text{Minimize} \quad f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_j) \\ & \text{subject to} \quad a_j \leq x_j \leq b_j, \end{aligned}$$

where $[a_j, b_j]$, with a_j, $b_j \geq 0$, is the interval of interest for x_j. Let $\overline{z}_j$ be the optimal objective function value to the above problem. Then the following hold true:

1. $\sum_{j \notin L} \overline{z}_j - \sum_{i=1}^{m} \hat{u}_i p_i \leq \sum_{j=1}^{n} f_j(\overline{x}_j) \leq \sum_{j=1}^{n} f_j(\hat{x}_j) \leq \hat{z}$, where $\hat{x}_j = \sum_{v=1}^{k_j} \hat{\lambda}_{vj} x_{vj}$ for $j \notin L$, and $\overline{\mathbf{x}} = (\overline{x}_1, \ldots, \overline{x}_n)^t$ is an optimal solution to Problem P.
2. If $\overline{z}_j + \hat{v}_j \geq 0$ for $j \notin L$, then $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$ is an optimal solution to Problem P. Furthermore, $\sum_{j=1}^{n} f_j(\hat{x}_j) = \hat{z}$.
3. If $\overline{z}_j + \hat{v}_j < 0$ for some $j \notin L$, let $x_{\gamma j}$ be the optimal solution that yielded $\overline{z}_j < -\hat{v}_j$. Then, adding the grid point $x_{\gamma j}$ in defining Problem LAP will give a new approximating Problem LAP with a minimum objective function value not higher than $\hat{z}$.

Proof

Since $\hat{u}_i$ and $\hat{v}_j$ are the optimal Lagrangian multipliers associated with Problem LAP, the reader can verify that the following subset of the KKT conditions hold true:

$$c_j + \sum_{i=1}^{m} \hat{u}_i a_{ij} \geq 0 \qquad \text{for } j \in L.$$

Multiplying by $x_j \geq 0$, and noting that $f_j(x_j) = c_j x_j$ and $g_{ij}(x_j) = a_{ij} x_j$, we get

$$f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_j) \geq 0 \qquad \text{for } j \in L \text{ and for all } x_j \geq 0. \tag{11.38}$$

Furthermore, from the definition of $\overline{z}_j$, we have

$$f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_j) \geq \overline{z}_j \qquad \text{for } j \notin L \text{ and for all } a_j \leq x_j \leq b_j. \tag{11.39}$$

Summing (11.38) over $j \in L$ and (11.39) over $j \notin L$, and subtracting $\sum_{i=1}^{m} \hat{u}_i p_i$ from the resulting sum, we get

$$\sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i \left[\sum_{j=1}^{n} g_{ij}(x_j) - p_i \right] \geq \sum_{j \notin L} \overline{z}_j - \sum_{i=1}^{m} \hat{u}_i p_i \tag{11.40}$$
$$\text{for all } a_j \leq x_j \leq b_j.$$

Noting that $a_j \le \bar{x}_j \le b_j$, $\sum_{j=1}^{n} g_{ij}(\bar{x}_j) \le p_i$, and that $\hat{u}_i \ge 0$, (11.40) implies that $\sum_{j=1}^{n} f_j(\bar{x}_j) \ge \sum_{j \notin L} \bar{z}_j - \sum_{i=1}^{m} \hat{u}_i p_i$, which is the first inequality in Part 1 of the theorem. Now, by Theorem 11.3.4, $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)^t$ is feasible to Problem P, so that $\sum_{j=1}^{n} f_j(\bar{x}_j) \le \sum_{j=1}^{n} f_j(\hat{x}_j)$. Finally, by the convexity of f_j for $j \notin L$, we have

$$\begin{aligned}
\sum_{j=1}^{m} f_j(\hat{x}_j) &= \sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} f_j(\hat{x}_j) \\
&= \sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} f_j\left[\sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} x_{\nu j}\right] \\
&\le \sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} f_j(x_{\nu j}) \\
&= \hat{z}.
\end{aligned}$$

Hence, Part 1 of the theorem holds true.

To prove Part 2, consider Problem LAP defined in (11.24). The reader can verify that the complementary slackness conditions of the KKT optimality conditions provide

$$f_j(\hat{x}_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(\hat{x}_j) = 0 \qquad \text{for } j \in L \tag{11.41}$$

$$\hat{\lambda}_{\nu j}\left[f_j(x_{\nu j}) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_{\nu j}) + \hat{v}_j\right] = 0 \qquad \text{for } \nu = 1, \ldots k_j, j \notin L \tag{11.42}$$

$$\hat{u}_i\left[\sum_{j \in L} g_{ij}(\hat{x}_j) + \sum_{j \notin L} \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} g_{ij}(x_{\nu j}) - p_i\right] = 0 \qquad \text{for } i = 1, \ldots, m. \tag{11.43}$$

Summing (11.41) over $j \in L$ and (11.42) over $\nu = 1, \ldots, k_j, j \notin L$, we get

$$\begin{aligned}
&\left[\sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} f_j(x_{\nu j})\right] + \sum_{i=1}^{m} \hat{u}_i \left[\sum_{j \in L} g_{ij}(\hat{x}_j) + \right. \\
&\qquad \left. + \sum_{j \notin L} \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} g_{ij}(x_{\nu j})\right] + \sum_{j \notin L} \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} \hat{v}_j = 0.
\end{aligned} \tag{11.44}$$

But the first term in (11.44) is precisely $\hat{z}$ by definition, and the second term is equal to $\sum_{i=1}^{m} \hat{u}_i p_i$ by (11.43). Furthermore, $\sum_{v=1}^{k_j} \hat{\lambda}_{vj} = 1$ for $j \notin L$, since $\hat{\lambda}_{vj}$ is feasible to Problem LAP defined in (11.24). Hence,

$$\hat{z} + \sum_{i=1}^{m} \hat{u}_i p_i + \sum_{j \notin L} \hat{v}_j = 0. \tag{11.45}$$

Also, from Part 1 of the theorem, we have

$$\sum_{j \notin L} \overline{z}_j - \sum_{i=1}^{m} \hat{u}_i p_i \le \sum_{j=1}^{n} f_j(\overline{x}_j). \tag{11.46}$$

Adding (11.45) to (11.46), we get $\sum_{j \notin L}(\overline{z}_j + \hat{v}_j) + \hat{z} \le \sum_{j=1}^{n} f_j(\overline{x}_j)$. But by assumption in Part 2, $\overline{z}_j + \hat{v}_j \ge 0$ for $j \notin L$. Hence, $\hat{z} \le \sum_{j=1}^{n} f_j(\overline{x}_j)$; and, using Part 1 of the theorem, we get, $\hat{z} \le \sum_{j=1}^{n} f_j(\overline{x}_j) \le \sum_{j=1}^{n} f_j(\hat{x}_j) \le \hat{z}$. This implies that $\sum_{j=1}^{n} f_j(\overline{x}_j) = \sum_{j=1}^{n} f_j(\hat{x}_j)$. Since $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)^t$ is feasible to Problem P, Part 2 follows.

To prove Part 3, suppose that $x_{\gamma j}$ is the optimal solution that yielded $\overline{z}_j < -v_j$. We then have $f_j(x_{\gamma j}) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_{\gamma j}) + \hat{v}_j < 0$. But if the grid point $x_{\gamma j}$ were included in defining the approximating Problem LAP, then one of the KKT conditions, namely, $f_j(x_{\gamma j}) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_{\gamma j}) + \hat{v}_j \ge 0$, would be violated. The reader can easily verify that introducing $x_{\gamma j}$ in the basis will yield an objective value in Problem LAP not higher than $\hat{z}$, and the proof is complete.

Summary of the Grid Point Generation Procedure

The procedure described below can be used to solve a problem of the form to minimize $\sum_{j=1}^{n} f_j(x_j)$ subject to $\sum_{j=1}^{n} g_{ij}(x_j) \le 0$ for $i = 1,\ldots, m$ and $x_j \ge 0$ for $j = 1,\ldots, n$. Let $L = \{j : f_j$ and g_{ij} for $i = 1,\ldots, m$ are linear$\}$. The procedure will yield an optimal solution using the simplex method without the restricted basis entry if g_{ij} is convex for $i = 1,\ldots, m$ and $j \notin L$, and f_j is strictly convex for $j \notin L$.

Initialization Step Define $a_j, b_j \ge 0$ such that all feasible points satisfy $x_j \in [a_j, b_j]$ for $j \notin L$. For each $j \notin L$, select a set of grid points. Set k_j equal to the number of grid points for $j \notin L$, and go to the Main Step.

Main Step

1. Solve Problem LAP defined in (11.24). Let the optimal solution be $\hat{x}_j$ for $j \in L$, and $\hat{\lambda}_{vj}$ for $v = 1,\ldots,k_j$, $j \notin L$. Let $\hat{u}_i$ be the Lagrangian multipliers associated with the first m constraints, and let $\hat{v}_j$ for $j \notin L$ be the Lagrangian multipliers associated with $\sum_{v=1}^{k_j} \lambda_{vj} = 1$. Go to Step 2.
2. For each $j \notin L$, solve the problem to minimize $f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_j)$, subject to $a_j \le x_j \le b_j$. Let the optimum objective function value be $\bar{z}_j$ for $j \notin L$. If $\bar{z}_j + \hat{v}_j \ge 0$ for all $j \notin L$, stop; the optimal solution to the original problem is $\hat{\mathbf{x}}$, whose components are given by $\hat{x}_j$ for $j \in L$ and $\hat{x}_j = \sum_{v=1}^{k_j} \hat{\lambda}_{vj} x_{vj}$. Otherwise, go to Step 3.
3. Let $\bar{z}_p + \hat{v}_p = \text{minimum}_{j \notin L}(\bar{z}_j + \hat{v}_j) < 0$. Let x_{vp} be the optimum solution yielding $\bar{z}_p < -\hat{v}_p$. Let $v = k_p + 1$, replace k_p by $k_p + 1$, and go to Step 1.

11.3.6 Example

Consider the following separable program:

$$
\begin{aligned}
\text{Minimize} \quad & x_1^2 - 6x_1 + x_2^2 - 8x_2 - \frac{1}{2}x_3 \\
\text{subject to} \quad & x_1 + x_2 + x_3 \le 5 \\
& x_1^2 - x_2 \le 3 \\
& x_1, x_2, x_3 \ge 0.
\end{aligned}
$$

Iteration **1:**

Since the objective and constraint functions associated with x_3 are linear, we let $L = \{3\}$. We start the grid generation procedure with the initial grid points $x_{11} = x_{12} = 0$. The corresponding columns are $(0, 0, 1, 0)^t$ and $(0, 0, 0, 1)^t$, and the corresponding objective values are both equal to zero. Letting x_4 and x_5 be the slack variables, we get the first tableau given below. At this stage, x_3 enters the basis and x_4 leaves the basis, giving the second tableau.

	z	λ_{11}	λ_{12}	x_3	x_4	x_5	RHS
z	1	0	0	0.5	0	0	0
x_4	0	0	0	(1)	1	0	5
x_5	0	0	0	0	0	1	3
λ_{11}	0	1	0	0	0	0	1
λ_{12}	0	0	1	0	0	0	1

	z	λ_{11}	λ_{12}	x_3	x_4	x_5	RHS
z	1	0	0	0	−0.5	0	−2.5
x_3	0	0	0	1	1	0	5
x_5	0	0	0	0	0	1	3
λ_{11}	0	1	0	0	0	0	1
λ_{12}	0	0	1	0	0	0	1

Note that $\hat{x}_j = \sum_v \hat{\lambda}_{vj} x_{vj}$ for $j = 1, 2$. From the second tableau, $\hat{\lambda}_{11} = \hat{\lambda}_{12} = 1$, so that $\hat{x}_1 = \hat{x}_2 = 0$. Therefore, the current solution $\hat{\mathbf{x}} = (0, 0, 5)^t$ and $f(\hat{\mathbf{x}}) = -2.5$. Note that the Lagrangian multipliers $\hat{u}_1$ and $\hat{u}_2$ associated with the constraints $x_1 + x_2 + x_3 \le 5$ and $x_1^2 - x_2 \le 3$ are the negatives of the entries in row 0 and under x_4 and x_5, so that $\hat{u}_1 = 0.5$ and $\hat{u}_2 = 0$. The Lagrangian multipliers $\hat{v}_1$ and $\hat{v}_2$ associated with the constraints $\sum_v \lambda_{v1} = 1$ and $\sum_v \lambda_{v2} = 1$ are the negatives of the entries in row 0 under λ_{11} and λ_{12}, so that $\hat{v}_1 = \hat{v}_2 = 0$. To find whether a new grid point is needed, we solve the following two problems:

$$\text{Minimize} \quad f_1(x_1) + \sum_{i=1}^{2} \hat{u}_i g_{i1}(x_1) = x_1^2 - 5.5x_1 \quad \text{subject to } 0 \le x_1 \le 5.$$

$$\text{Minimize} \quad f_2(x_2) + \sum_{i=1}^{2} \hat{u}_i g_{i2}(x_2) = x_2^2 - 7.5x_2 \quad \text{subject to } 0 \le x_2 \le 5.$$

For the first problem, the optimal solution is $\bar{x}_1 = 2.75$ with optimal objective value $\bar{z}_1 = -7.56$. Thus, $\bar{z}_1 + \hat{v}_1 = -7.56 < 0$, and the grid point $\bar{x}_1 = 2.75$ would improve the objective value if introduced. For the second problem, the optimal solution is $\bar{x}_2 = 3.75$ with optimal objective value $\bar{z}_2 = -14.06$. Thus, $\bar{z}_2 + \hat{v}_2 = -14.06 < 0$, and the grid point $\bar{x}_2 = 3.75$ would also improve the objective function if introduced. Since minimum $\{\bar{z}_1 + \hat{v}_1, \bar{z}_2 + \hat{v}_2\} = \bar{z}_2 + \hat{v}_2 = -14.06$, we introduce the grid point $x_{22} = \bar{x}_2 = 3.75$. The variable associated with the grid point x_{22} is λ_{22}. (Computationally, $\bar{x}_1$ may be stored temporarily and entered sequentially if it remains enterable following the chosen pivot operation.)

Iteration 2:

Note that $g_{12}(x_{22}) = 3.75$ and $g_{22}(x_{22}) = -3.75$, so that the column associated with x_{22} is $(3.75, -3.75, 0, 1)^t$. This column needs to be updated by premultiplying it by the basis inverse $\mathbf{B}^{-1}$. From the last tableau, $\mathbf{B}^{-1} = \mathbf{I}$, and hence, the updated column for λ_{22} is $(3.75, -3.75, 0, 1)^t$. The updated coefficient in row 0 is given by $-(\bar{z}_2 + \hat{v}_2) = 14.06$. The associated tableau is given below, and λ_{22} enters the basis giving the second tableau.

	z	λ_{11}	λ_{12}	λ_{22}	x_3	x_4	x_5	RHS
z	1	0	0	14.06	0	−0.5	0	−2.5
x_3	0	0	0	3.75	1	1	0	5
x_5	0	0	0	−3.75	0	0	1	3
λ_{11}	0	1	0	0	0	0	0	1
λ_{12}	0	0	1	(1)	0	0	0	1

	z	λ_{11}	λ_{12}	λ_{22}	x_3	x_4	x_5	RHS
z	1	0	−14.06	0	0	−0.5	0	−16.56
x_3	0	0	−3.75	0	1	1	0	1.25
x_5	0	0	3.75	0	0	0	1	6.75
λ_{11}	0	1	0	0	0	0	0	1
λ_{22}	0	0	1	1	0	0	0	1

From the last tableau, $\hat{\lambda}_{11} = \hat{\lambda}_{22} = 1$ and $\hat{\lambda}_{12} = 0$. Noting that $\hat{x}_j = \sum_v \hat{\lambda}_{vj} x_{vj}$ for $j = 1, 2$ it follows that $\hat{x}_1 = 0$ and $\hat{x}_2 = 3.75$. Since $\hat{x}_3 = 1.25$, the current solution is $\hat{\mathbf{x}} = (0, 3.75, 1.25)^t$ and $f(\hat{\mathbf{x}}) = -17.19$. From the above tableau, $\hat{u}_1 = 0.5$, $\hat{u}_2 = 0$, $\hat{v}_1 = 0$, and $\hat{v}_2 = 14.06$. Since the values of $\hat{u}_1$ and $\hat{u}_2$ did not change from those at Iteration 1, $\bar{x}_1 = 2.75$ and $\bar{x}_2 = 3.75$ remain optimal. Note that $\bar{z}_1 = -7.56$ and $\bar{z}_2 = -14.06$, so that $\min\{\bar{z}_1 + \hat{v}_1, \bar{z}_2 + \hat{v}_2\} = \bar{z}_1 + \hat{v}_1 = -7.56$. Thus, we introduce the grid point $x_{21} = \bar{x}_1 = 2.75$. The variable corresponding to x_{21} is λ_{21}.

Iteration 3:

Note that $g_{11}(x_{21}) = 2.75$ and $g_{21}(x_{21}) = 7.56$, so that the column associated with x_{21} is $(2.75, 7.56, 1, 0)^t$. From the last tableau, the basis inverse $\mathbf{B}^{-1}$ is given by

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 & -3.75 \\ 0 & 1 & 0 & 3.75 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Hence, the updated column for λ_{21} is $\mathbf{B}^{-1}(2.75, 7.56, 1, 0)' = (2.75, 7.56, 1, 0)'$. The entry in row 0 under λ_{21} is given by $-(\bar{z}_1 + \hat{v}_1) = 7.56$. The associated tableau is given below, and λ_{21} enters the basis giving the second tableau.

	z	λ_{11}	λ_{21}	λ_{12}	λ_{22}	x_3	x_4	x_5	RHS
z	1	0	7.56	−14.06	0	0	−0.5	0	−16.56
x_3	0	0	(2.75)	−3.75	0	1	1	0	1.25
x_5	0	0	7.56	3.75	0	0	0	1	6.75
λ_{11}	0	1	1	0	0	0	0	0	1
λ_{12}	0	0	0	1	1	0	0	0	1

	z	λ_{11}	λ_{21}	λ_{12}	λ_{22}	x_3	x_4	x_5	RHS
z	1	0	0	−3.78	0	−2.72	−3.22	0	−19.96
λ_{21}	0	0	1	−1.36	0	0.36	0.36	0	0.45
x_5	0	0	0	14.03	0	−2.72	−2.72	1	3.35
λ_{11}	0	1	0	1.36	0	−0.36	−0.36	0	0.55
λ_{12}	0	0	0	1	1	0	0	0	1

From the above tableau, $\hat{\lambda}_{11} = 0.55$, $\hat{\lambda}_{21} = 0.45$, $\hat{\lambda}_{12} = 0$, and $\hat{\lambda}_{22} = 1$. Therefore, $\hat{x}_1 = 1.25$ and $\hat{x}_2 = 3.75$. The current solution is thus $\hat{\mathbf{x}} = (1.25, 3.75, 0)'$ and $f(\hat{\mathbf{x}}) = -21.88$. From the last tableau, $\hat{u}_1 = 3.22$, $\hat{u}_2 = 0$, $\hat{v}_1 = 0$, and $\hat{v}_2 = 3.78$. To find whether a new point is needed, we solve the following two problems:

$$\text{Minimize} \quad f_1(x_1) + \sum_{i=1}^{2} \hat{u}_i g_{i1}(x_1) = x_1^2 - 2.78x_1 \quad \text{subject to } 0 \le x_1 \le 5.$$

$$\text{Minimize} \quad f_2(x_2) + \sum_{i=1}^{2} \hat{u}_i g_{i2}(x_2) = x_2^2 - 4.78x_2 \quad \text{subject to } 0 \le x_2 \le 5.$$

The optimal solution to the first problem is $\bar{x}_1 = 1.39$ and the optimal objective value is $\bar{z}_1 = -1.93$. The optimal solution to the second problem is $\bar{x}_2 = 2.39$ and the optimal objective value is $\bar{z}_2 = -5.71$. Thus, $\min\{\bar{z}_1 + \hat{v}_1, \bar{z}_2 + \hat{v}_2\} = \bar{z}_1 + \hat{v}_1 =$

$\bar{z}_2 + \hat{v}_2 = -1.93$. Therefore, we can introduce either the grid point $\bar{x}_1 = 1.39$ or the grid point $\bar{x}_2 = 2.39$. Note that

$$\sum_{j=1}^{2} \bar{z}_j - \sum_{i=1}^{2} \hat{u}_i p_i = -23.74 \qquad \text{and} \qquad f(\hat{\mathbf{x}}) = -21.88.$$

By Part 1 of Theorem 11.3.5, the optimal objective value of the original problem lies between −23.74 and −21.88. Thus, if we stop the algorithm at this stage, we would have a feasible solution $\hat{\mathbf{x}} = (1.25, 3.75, 0)^t$ whose objective value is −21.88, and we would also know that a lower bound on the optimal objective value to the original problem is −23.74. If more accuracy is desired, the process would continue by introducing the new grid point $x_{31} = 1.39$ or the new grid point $x_{32} = 2.39$.

11.4 Linear Fractional Programming

In this section we consider a problem in which the objective function is the ratio of two linear functions and the constraints are linear. Such problems are called *linear fractional programming problems* and can be stated precisely as follows:

$$\begin{aligned} \text{Minimize} \quad & \frac{\mathbf{p}^t\mathbf{x}+\alpha}{\mathbf{q}^t\mathbf{x}+\beta} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x}=\mathbf{b} \\ & \mathbf{x}\geq \mathbf{0}, \end{aligned}$$

where $\mathbf{p}$ and $\mathbf{q}$ are n-vectors, $\mathbf{b}$ is an m-vector, $\mathbf{A}$ is an $m \times n$ matrix, and α and β are scalars. As we shall soon observe, if an optimal solution for a linear fractional program exists, then an extreme point optimum exists. Furthermore, every local minimum is a global minimum. Hence, a procedure that moves from one extreme point to an adjacent one is a viable approach for solving such a problem. Lemma 11.4.1 gives some important properties of the objective function.

11.4.1 Lemma

Let $f(\mathbf{x}) = (\mathbf{p}^t\mathbf{x}+\alpha)/(\mathbf{q}^t\mathbf{x}+\beta)$, and let S be a convex set such that $\mathbf{q}^t\mathbf{x} + \beta \neq 0$ over S. Then f is both pseudoconvex and pseudoconcave over S.

Proof

First, note that either $\mathbf{q}^t\mathbf{x} + \beta > 0$ for all $\mathbf{x} \in S$ or $\mathbf{q}^t\mathbf{x} + \beta < 0$ for all $\mathbf{x} \in S$. Otherwise, there exist $\mathbf{x}_1$ and $\mathbf{x}_2$ in S such that $\mathbf{q}^t\mathbf{x}_1 + \beta > 0$ and $\mathbf{q}^t\mathbf{x}_2 + \beta < 0$; and hence for some convex combination $\mathbf{x}$ of $\mathbf{x}_1$ and $\mathbf{x}_2$, $\mathbf{q}^t\mathbf{x} + \beta = 0$, contradicting our assumption. We first show that f is pseudoconvex. Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in S$ with $(\mathbf{x}_2 - \mathbf{x}_1)^t \nabla f(\mathbf{x}_1) \ge 0$. We need to show that $f(\mathbf{x}_2) \ge f(\mathbf{x}_1)$. Note that

$$\nabla f(\mathbf{x}_1) = \frac{(\mathbf{q}^t\mathbf{x}_1 + \beta)\mathbf{p} - (\mathbf{p}^t\mathbf{x}_1 + \alpha)\mathbf{q}}{(\mathbf{q}^t\mathbf{x}_1 + \beta)^2}.$$

Since $(\mathbf{x}_2 - \mathbf{x}_1)^t \nabla f(\mathbf{x}_1) \ge 0$ and since $(\mathbf{q}^t\mathbf{x}_1 + \beta)^2 > 0$, it follows that

$$\begin{aligned} 0 &\le (\mathbf{x}_2 - \mathbf{x}_1)^t[(\mathbf{q}^t\mathbf{x}_1 + \beta)\mathbf{p} - (\mathbf{p}^t\mathbf{x}_1 + \alpha)\mathbf{q}] \\ &= (\mathbf{p}^t\mathbf{x}_2 + \alpha)(\mathbf{q}^t\mathbf{x}_1 + \beta) - (\mathbf{q}^t\mathbf{x}_2 + \beta)(\mathbf{p}^t\mathbf{x}_1 + \alpha). \end{aligned}$$

Therefore, $(\mathbf{p}^t\mathbf{x}_2 + \alpha)(\mathbf{q}^t\mathbf{x}_1 + \beta) \ge (\mathbf{q}^t\mathbf{x}_2 + \beta)(\mathbf{p}^t\mathbf{x}_1 + \alpha)$. But since $\mathbf{q}^t\mathbf{x}_1 + \beta$ and $\mathbf{q}^t\mathbf{x}_2 + \beta$ are both either positive or negative, dividing by $(\mathbf{q}^t\mathbf{x}_1 + \beta)(\mathbf{q}^t\mathbf{x}_2 + \beta) > 0$, we get

$$\frac{\mathbf{p}^t\mathbf{x}_2 + \alpha}{\mathbf{q}^t\mathbf{x}_2 + \beta} \ge \frac{\mathbf{p}^t\mathbf{x}_1 + \alpha}{\mathbf{q}^t\mathbf{x}_1 + \beta}; \qquad \text{that is, } f(\mathbf{x}_2) \ge f(\mathbf{x}_1).$$

Therefore, f is pseudoconvex. Similarly, it can be shown that $(\mathbf{x}_2 - \mathbf{x}_1)^t \nabla f(\mathbf{x}_1) \le 0$ implies that $f(\mathbf{x}_2) \le f(\mathbf{x}_1)$, and hence, f is pseudoconcave, and the proof is complete.

Several implications of Lemma 11.4.1 for a linear fractional programming problem may be noted.

1. Since the objective function is both pseudoconvex and pseudoconcave over S, then by Theorem 3.5.11, it is also quasiconvex, quasiconcave, strictly quasiconvex, and strictly quasiconcave.
2. Since the objective function is both pseudoconvex and pseudoconcave, then, by Theorem 4.3.8, a point satisfying the KKT conditions for a minimization problem is also a global minimum over the feasible region. Similarly, a point satisfying the KKT conditions for a maximization problem is also a global maximum over the feasible region.
3. Since the objective function is strictly quasiconvex and strictly quasiconcave, then, by Theorem 3.5.6, a local minimum is also a global minimum over the feasible region. Similarly, a local maximum is also a global maximum over the feasible region.
4. Since the objective function is quasiconcave and quasiconvex, if the feasible region is bounded, then, by Theorem 3.5.3, the objective function has a minimum at an extreme point of the feasible region and also has a maximum at an extreme point of the feasible region.

The foregoing facts about the objective function f give very useful results that can be used to develop suitable computational procedures for solving the fractional programming problem. In particular, we may search among the extreme points of the polyhedral set $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ until a KKT point is reached. We now show that the convex-simplex method gives a convenient solution procedure.

Minimization by the Convex-Simplex Method

Because of the special structure of the objective function f, the convex-simplex method simplifies into a minor modification of the simplex method of linear programming. Suppose that we are given an extreme point of the feasible region with basis $\mathbf{B}$ such that $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$ and $\mathbf{x}_N = \mathbf{0}$. Recall from Section 10.7 that the convex-simplex method increases or decreases one of the nonbasic variables and then modifies the basic variables accordingly. Since the current point is an extreme point with $\mathbf{x}_N = \mathbf{0}$, decreasing a nonbasic variable is not permitted, as it would violate the nonnegativity restriction. Thus, the direction-finding process simplifies as follows. Let $\mathbf{r}_N$ denote the nonbasic components of the reduced gradient vector $\mathbf{r}^t = \nabla f(\mathbf{x})^t - \nabla_B f(\mathbf{x})^t \mathbf{B}^{-1}\mathbf{A}$, so that

$$\mathbf{r}_N^t = \nabla_N f(\mathbf{x})^t + \nabla_B f(\mathbf{x})^t \mathbf{B}^{-1}\mathbf{N}.$$

By Theorem 10.5.1, if $\mathbf{r}_N \geq \mathbf{0}$, then the current point is a KKT point, and we must stop. Otherwise, let $-r_j = \max\{-r_i : r_i \leq 0\}$, where r_i is the ith component of $\mathbf{r}_N$. The nonbasic variable x_j is increased, and the basic variables are modified to maintain feasibility. This is equivalent to moving along the direction d, whose nonbasic and basic components $\mathbf{d}_N$ and $\mathbf{d}_B$ are given as follows. The direction $\mathbf{d}_N$ is a vector of zeros, except for a 1 at the jth position, and $\mathbf{d}_B = -\mathbf{B}^{-1}\mathbf{a}_j$, where $\mathbf{a}_j$ is the jth column of $\mathbf{A}$. By Theorem 10.6.1, $\mathbf{d}$ is an improving feasible direction. As we shall see by Lemma 11.4.2, no line search along the direction $\mathbf{d}$ is needed. Indeed, due to the special structure of the objective function, if $\nabla f(\mathbf{x})^t\mathbf{d} < 0$, then the function f continues to decrease by moving along $\mathbf{d}$. Thus, we move along $\mathbf{d}$ as far as possible. Since moving along $\mathbf{d}$ is equivalent to increasing a nonbasic variable and adjusting the basic variables, we move along $\mathbf{d}$ until a basic variable drops to zero and leaves the basis, producing an adjacent extreme point. The entire process is then repeated.

11.4.2 Lemma

Let $f(\mathbf{x}) = (\mathbf{p}^t\mathbf{x} + \alpha)/(\mathbf{q}^t\mathbf{x} + \beta)$, and let S be a convex set. Furthermore, suppose that $\mathbf{q}^t\mathbf{x} + \beta \neq 0$ on S. Given $\mathbf{x} \in S$, let $\mathbf{d}$ be such that $\nabla f(\mathbf{x})^t\mathbf{d} < 0$. Then $f(\mathbf{x} + \lambda\mathbf{d})$ is a decreasing function of λ.

Proof

Note that

$$\nabla f(\mathbf{y}) = \frac{(\mathbf{q}^t\mathbf{y}+\beta)\mathbf{p}-(\mathbf{p}^t\mathbf{y}+\alpha)\mathbf{q}}{(\mathbf{q}^t\mathbf{y}+\beta)^2}. \tag{11.47}$$

Letting $\mathbf{y} = \mathbf{x} + \lambda\mathbf{d}$, $s = [\mathbf{q}^t(\mathbf{x}+\lambda\mathbf{d})+\beta]^2 > 0$, and $s' = (\mathbf{q}^t\mathbf{x}+\beta)^2 > 0$, we get

$$\begin{aligned}\nabla f(\mathbf{x}+\lambda\mathbf{d}) &= \frac{[\mathbf{q}^t(\mathbf{x}+\lambda\mathbf{d})+\beta]\mathbf{p}-[\mathbf{p}^t(\mathbf{x}+\lambda\mathbf{d})+\alpha]\mathbf{q}}{s} \\ &= \frac{s'}{s}\nabla f(\mathbf{x})+\frac{\lambda}{s}[(\mathbf{q}^t\mathbf{d})\mathbf{p}-(\mathbf{p}^t\mathbf{d})\mathbf{q}].\end{aligned}$$

Therefore,

$$\begin{aligned}\nabla f(\mathbf{x}+\lambda\mathbf{d})^t\mathbf{d} &= \frac{s'}{s}\nabla f(\mathbf{x})^t\mathbf{d}+\frac{\lambda}{s}[(\mathbf{q}^t\mathbf{d})(\mathbf{p}^t\mathbf{d})-(\mathbf{p}^t\mathbf{d})(\mathbf{q}^t\mathbf{d})] \\ &= \frac{s'}{s}\nabla f(\mathbf{x})^t\mathbf{d}.\end{aligned} \tag{11.48}$$

Now let $\theta(\lambda) = f(\mathbf{x}+\lambda\mathbf{d})$. Then, by (11.48), $\theta'(\lambda) = \nabla f(\mathbf{x}+\lambda\mathbf{d})^t\mathbf{d} < 0$ for all λ, and the result follows.

To summarize, given the extreme point $\mathbf{x}$ and the direction $\mathbf{d}$ with $\nabla f(\mathbf{x})^t\mathbf{d} < 0$ as above, no minimization of f along $\mathbf{d}$ is necessary, since $f(\mathbf{x} + \lambda\mathbf{d})$ is a decreasing function of λ. Therefore, we move along $\mathbf{d}$ as much as possible, that is, until an adjacent extreme point is reached, and we then repeat the process. A precise summary of the algorithm utilizing a tableau format for updating the extreme points generated is presented below.

Summary of the Fractional Programming Algorithm of Gilmore and Gomory

We present below a method credited to Gilmore and Gomory [1963] for solving a linear fractional program of the form to minimize $(\mathbf{p}^t\mathbf{x}+\alpha)/(\mathbf{q}^t\mathbf{x}+\beta)$ subject to $\mathbf{x} \in S = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$. We will assume that the set S is bounded and that $\mathbf{q}^t\mathbf{x}+\beta \neq 0$ for all $\mathbf{x} \in S$.

Initialization Step Find a starting basic feasible solution $\mathbf{x}_1$ to the system $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$. Form the corresponding tableau represented by $\mathbf{x}_B + \mathbf{B}^{-1}\mathbf{N}\mathbf{x}_N = \mathbf{B}^{-1}\mathbf{b}$. Let $k = 1$ and go to the Main Step.

Main Step

1. Compute the vector $\mathbf{r}_N^t = \nabla_N f(\mathbf{x}_k)^t - \nabla_B f(\mathbf{x}_k)^t \mathbf{B}^{-1}\mathbf{N}$. If $\mathbf{r}_N \geq 0$, stop; the current point $\mathbf{x}_k$ is an optimal solution. Otherwise, go to Step 2.
2. Let $-r_j = \max\{-r_i : r_i \leq 0\}$, where r_i is the ith component of $\mathbf{r}_N$. Determine the basic variable $\mathbf{x}_{B_r}$ to leave the basis by the following minimum ratio test:

$$\frac{\bar{b}_r}{y_{rj}} = \min_{1 \leq i \leq m} \left\{ \frac{\bar{b}_i}{y_{ij}} : y_{ij} > 0 \right\},$$

   where $\bar{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b}$, $\mathbf{y}_j = \mathbf{B}^{-1}\mathbf{a}_j$, and $\mathbf{a}_j$ is the jth column of $\mathbf{A}$. Go to Step 3.
3. Replace the variable x_{B_r} by the variable x_j. Update the tableau correspondingly by pivoting at y_{rj}. Let the current solution be $\mathbf{x}_{k+1}$. Replace k by $k + 1$ and go to Step 1.

Exercise 11.43 shows that the reduced gradient $\mathbf{r}_N$ could readily be computed if two additional rows, one corresponding to $\mathbf{p}^t\mathbf{x} + \alpha$ and the other corresponding to $\mathbf{q}^t\mathbf{x} + \beta$, are introduced and carried forward at each iteration.

Finite Convergence

We now establish finite convergence under the nondegeneracy assumption that $\mathbf{x}_B > \mathbf{0}$ for each extreme point. Note that the algorithm moves from one extreme point to another. By Lemma 11.4.2 and the nondegeneracy assumption, the objective function strictly decreases at each iteration so that the extreme points generated are distinct. There exists only a finite number of these points, and hence, the algorithm stops in a finite number of steps. At termination, the reduced gradient is nonnegative resulting in a KKT point; and by Lemma 11.4.1, this point is indeed an optimal point.

11.4.3 Example

Consider the following linear fractional program:

$$\text{Minimize} \quad \frac{-2x_1 + x_2 + 2}{x_1 + 3x_2 + 4}$$
$$\text{subject to} \quad -x_1 + x_2 \le 4$$
$$x_2 \le 6$$
$$2x_1 + x_2 \le 14$$
$$x_1, x_2 \ge 0.$$

Figure 11.5 depicts the feasible region with the extreme points (0, 0), (0, 4), (2, 6), (4, 6), and (7, 0). The objective values at these points are 0.5, 0.375, 0.167, 0.0, and −1.09, respectively, and hence, the optimal point is (7, 0).

Introducing the slack variables x_3, x_4, and x_5, we get the initial extreme point $\mathbf{x}_1 = (0, 0, 4, 6, 14)^t$.

Iteration 1:

The following tableau summarizes the computations for this iteration.

	x_1	x_2	x_3	x_4	x_5	RHS
$\nabla f(\mathbf{x}_1)$	−10/16	−2/16	0	0	0	—
x_3	−1	1	1	0	0	4
x_4	0	1	0	1	0	6
x_5	2	1	0	0	1	14
r	−10/16	−2/16	0	0	0	—

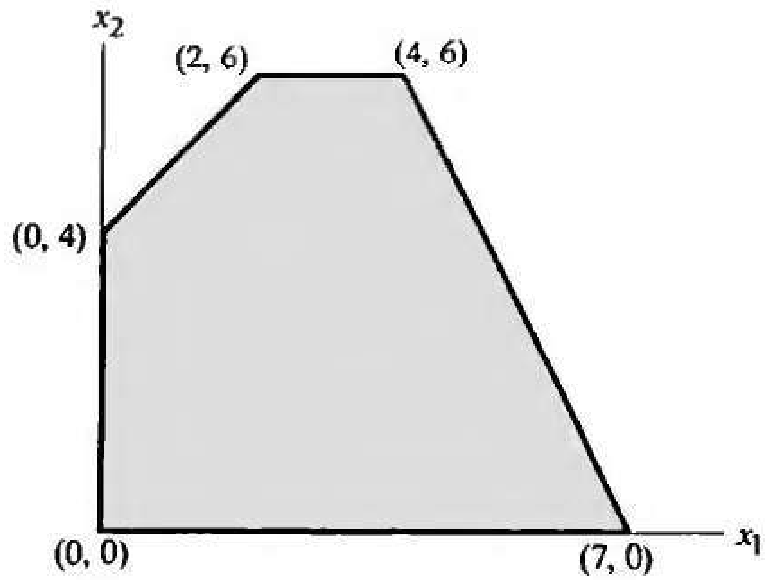


Figure 11.5 Feasible region for Example 11.4.3.

We have $\mathbf{q}^t\mathbf{x}_1 + \beta = 4$ and $\mathbf{p}^t\mathbf{x}_1 + \alpha = 2$. Hence, from (11.47), we get $\nabla f(\mathbf{x})^t = (-10/16, -2/16, 0, 0, 0)$, $\nabla_N f(\mathbf{x})^t = (-10/16, -2/16)$, and $\nabla_B f(\mathbf{x})^t = (0, 0, 0)$. The columns of x_1 and x_2 give $\mathbf{B}^{-1}\mathbf{N}$, and we get

$$\begin{aligned}\mathbf{r}_N^t = (r_1, r_2) &= \nabla_N f(\mathbf{x}_1)^t - \nabla_B f(\mathbf{x}_1)^t \mathbf{B}^{-1}\mathbf{N} \\ &= \left(-\frac{10}{16}, -\frac{2}{16}\right) - (0,0,0)\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} = \left(-\frac{10}{16}, -\frac{2}{16}\right).\end{aligned}$$

Note that $\mathbf{r}_N^t = (r_3, r_4, r_5) = (0, 0, 0)$. Since $\max\{-r_1, -r_2, -r_3, -r_4, -r_5\} = 10/16$, x_1 enters the basis. By the minimum ratio test, x_5 leaves the basis.

Iteration 2:

The computations for this iteration are summarized below.

	x_1	x_2	x_3	x_4	x_5	RHS
$\nabla f(\mathbf{x}_2)$	−10/121	47/121	0	0	0	—
x_3	0	3/2	1	0	1/2	11
x_4	0	1	0	1	0	6
x_1	1	1/2	0	0	1/2	7
$\mathbf{r}$	0	52/121	0	0	5/121	—

When x_1 replaces x_5 in the basis, we get the point $\mathbf{x}_2^t = (7, 0, 11, 6, 0)$. Now, $\mathbf{q}^t\mathbf{x}_2 + \beta = 11$ and $\mathbf{p}^t\mathbf{x}_2 + \alpha = -12$, so that from (11.47) we get $\nabla f(\mathbf{x}_2)^t = (-10/121, 47/121, 0, 0, 0)$. Then, $\mathbf{B}^{-1}\mathbf{N}$ is given by the columns of x_2 and x_5 in the tableau, and we then get

$$\begin{aligned}\mathbf{r}_N^t = (r_2, r_5) &= \nabla_N f(\mathbf{x}_2)^t - \nabla_B f(\mathbf{x}_2)^t \mathbf{B}^{-1}\mathbf{N} \\ &= \left(\frac{47}{121}, 0\right) - \left(0, 0, -\frac{10}{121}\right)\begin{bmatrix} 3/2 & 1/2 \\ 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \\ &= \left(\frac{52}{121}, \frac{5}{121}\right).\end{aligned}$$

Since $\mathbf{r}_N \geq \mathbf{0}$, we stop with the optimal solution $x_1 = 7$ and $x_2 = 0$. The corresponding objective function value is −1.09.

Method of Charnes and Cooper [1962]

We now describe another procedure using the simplex method for solving a linear fractional programming problem. Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & \frac{\mathbf{p}^t\mathbf{x}+\alpha}{\mathbf{q}^t\mathbf{x}+\beta} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \ge \mathbf{0}. \end{aligned}$$

Suppose that the set $S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \le \mathbf{b} \text{ and } \mathbf{x} \ge \mathbf{0}\}$ is compact, and suppose that $\mathbf{q}^t\mathbf{x} + \beta > 0$ for each $\mathbf{x} \in S$. Letting $z = 1/(\mathbf{q}^t\mathbf{x} + \beta)$ and $\mathbf{y} = z\mathbf{x}$, and multiplying the constraints $\mathbf{A}\mathbf{x} \le \mathbf{b}$ by z, the above problem leads to the following linear program:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{p}^t\mathbf{y}+\alpha z \\ \text{subject to} \quad & \mathbf{A}\mathbf{y} - \mathbf{b}z \le \mathbf{0} \\ & \mathbf{q}^t\mathbf{y}+\beta z = 1 \\ & \mathbf{y} \ge \mathbf{0} \\ & z \ge 0. \end{aligned}$$

First, note that if $(\mathbf{y}, z)$ is a feasible solution to the above problem, then $z > 0$. This follows since if $z = 0$, then $\mathbf{y} \ne \mathbf{0}$ must be such that $\mathbf{A}\mathbf{y} \le \mathbf{0}$ and $\mathbf{y} \ge \mathbf{0}$, which means that $\mathbf{y}$ is a direction of S, violating the compactness assumption. We now demonstrate that if $(\bar{\mathbf{y}}, \bar{z})$ is an optimal solution to the above linear program, then $\bar{\mathbf{x}} = \bar{\mathbf{y}}/\bar{z}$ is an optimal solution to the fractional program.

Note that $\mathbf{A}\bar{\mathbf{x}} \le \mathbf{b}$ and $\bar{\mathbf{x}} \ge \mathbf{0}$, so that $\bar{\mathbf{x}}$ is a feasible solution to the fractional program. To show optimality of $\bar{\mathbf{x}}$, let $\mathbf{x}$ be such that $\mathbf{A}\mathbf{x} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$. Note that $\mathbf{q}^t\mathbf{x} + \beta > 0$ by assumption, and that the vector $(\mathbf{y}, z)$ is a feasible solution to the linear program, where $\mathbf{y} = \mathbf{x}/(\mathbf{q}^t\mathbf{x} + \beta)$ and $z = 1/(\mathbf{q}^t\mathbf{x} + \beta)$. Since $(\bar{\mathbf{y}}, \bar{z})$ is an optimal solution to the linear program, $\mathbf{p}^t\bar{\mathbf{y}} + \alpha\bar{z} \le \mathbf{p}^t\mathbf{y} + \alpha z$. Substituting for $\bar{\mathbf{y}}$, $\mathbf{y}$, and z, this inequality gives $\bar{z}(\mathbf{p}^t\bar{\mathbf{x}} + \alpha) \le (\mathbf{p}^t\mathbf{x}+\alpha)/(\mathbf{q}^t\mathbf{x} + \beta)$. The result follows immediately by dividing the left-hand side by $1 = \mathbf{q}^t\bar{\mathbf{y}} + \beta\bar{z}$.

Now if $\mathbf{q}^t\mathbf{x} + \beta < 0$ for all $\mathbf{x} \in S$, then letting $-z = 1/(\mathbf{q}^t\mathbf{x} + \beta)$ and $\mathbf{y} = z\mathbf{x}$ gives the following linear program:

$$\begin{aligned} \text{Minimize} \quad & -\mathbf{p}^t\mathbf{y} - \alpha z \\ \text{subject to} \quad & \mathbf{A}\mathbf{y} - \mathbf{b}z \le \mathbf{0} \\ & -\mathbf{q}^t\mathbf{y} - \beta z = 1 \\ & \mathbf{y} \ge \mathbf{0} \\ & z \ge 0. \end{aligned}$$

In a fashion similar to that above, if $(\bar{\mathbf{y}}, \bar{z})$ solves the above linear program, then $\bar{\mathbf{x}} = \bar{\mathbf{y}}/\bar{z}$ solves the fractional programming problem.

To summarize, we have shown that a fractional linear program can be solved via a linear programming problem having one additional variable and one additional constraint. The form of the linear program used depends on whether $\mathbf{q}^t\mathbf{x} + \beta > 0$ for all $\mathbf{x} \in S$ or $\mathbf{q}^t\mathbf{x} + \beta < 0$ for all $\mathbf{x} \in S$. If there exist $\mathbf{x}_1, \mathbf{x}_2 \in S$ such that $\mathbf{q}^t\mathbf{x}_1 + \beta > 0$ and $\mathbf{q}^t\mathbf{x}_2 + \beta < 0$, then the optimal solution to the fractional program is unbounded.

11.4.4 Example

Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & \frac{-2x_1 + x_2 + 2}{x_1 + 3x_2 + 4} \\ \text{subject to} \quad & -x_1 + x_2 \le 4 \\ & 2x_1 + x_2 \le 14 \\ & x_2 \le 6 \\ & x_1, x_2 \ge 0. \end{aligned}$$

The feasible region for this problem is shown in Figure 11.5. We solve this problem using the method of Charnes and Cooper. Note that the point (0, 0) is feasible and that at this point, $-x_1 + 3x_2 + 4 > 0$. Hence, the denominator is positive over the entire feasible region. The equivalent linear program is given by:

$$\begin{aligned} \text{Minimize} \quad & -2y_1 + y_2 + 2z \\ \text{subject to} \quad & -y_1 + y_2 - 4z \le 0 \\ & 2y_1 + y_2 - 14z \le 0 \\ & y_2 - 6z \le 0 \\ & y_1 + 3y_2 + 4z = 1 \\ & y_1, y_2, z \ge 0. \end{aligned}$$

The reader can verify that $y_1 = 7/11$, $y_2 = 0$, and $z = 1/11$ is an optimal solution to the above linear program. Hence, an optimal solution to the original problem is $x_1 = y_1/z_1 = 7$ and $x_2 = y_2/z_2 = 0$.

11.5 Geometric Programming

In this section we consider problems of the type

$$\begin{aligned} \text{GP: } & \text{Minimize} && f(\mathbf{x}) \\ & \text{subject to} && g_i(\mathbf{x}) \le 1 \quad \text{for } i = 1,\ldots,m \\ &&& \mathbf{x} > \mathbf{0}, \end{aligned}$$

where each of the functions f and g_i is a posynomial of $\mathbf{x} \in R^n$ and where the variables $\mathbf{x}$ assume strictly positive values by the nature of the problem itself. A *posynomial* is a function composed of terms of the type

$$T_k = \alpha_k \prod_{j=1}^{n} x_j^{a_{kj}} \tag{11.49}$$

where $\alpha_k > 0$ and where the exponents a_{kj}, $j = 1,\ldots, n$, are rational numbers that can be of either sign. In particular, for $\mathbf{x} > \mathbf{0}$, we have $T_k > 0$ as well. Hence, the objective and constraint functions can be written as

$$f(\mathbf{x}) = \sum_{k \in J_0} T_k, \text{ and } g_i(\mathbf{x}) = \sum_{k \in J_i} T_k \quad \text{for } i = 1,\ldots, m, \tag{11.50a}$$

where the collection of index sets $J_0, J_1,\ldots, J_m$ are mutually disjoint, and where

$$J_0 \cup J_1 \cup \cdots \cup J_m \equiv \{1, 2, \ldots, M\} \tag{11.50b}$$

represents a total of M terms, each of the type (11.49). Problems GP of this type are called *posynomial programming problems.* When the coefficients α_k are permitted to be negative, the functions (11.50a) are called *signomials*, and Problem GP is then known as a *signomial programming problem.* In either case, the problem is called a *geometric programming problem*, the name arising from the geometric-arithmetic mean inequality (see Exercise 4.15) used in the original analysis presented by Duffin et al. [1967] to transform the problem into a simpler equivalent form. As we shall see shortly, posynomial programming problems (which are of present concern) are convex programs. However, the more general signomial programming problems are nonconvex and require a different solution approach. For example, a reformulation–linearization technique (RLT) similar to the one described for solving quadratic programming problems in Section 11.2 can be designed to solve this class of problems to global optimality (see the Notes and References section).

Posynomial geometric programming problems arise frequently in engineering applications where the decision variables $\mathbf{x}$ are design variables that are required to take on positive values to be meaningful and where the objective and constraint functions model fundamental physical or economical relationships that, by their nature, turn out to be posynomials or may be transformed into such functions (see Exercises 11.51, 11.52, and 11.53). Ominous as this problem might appear, there exists a transformation that considerably simplifies it, often rendering it solvable as a linear system of equations, or as a manageable, linearly constrained problem. This transformation involves two steps of using a change of variables, interposed by an application of Lagrangian duality concepts from Chapter 6.

To introduce the first change of variables, let us substitute

$$y_j = \ln(x_j) \qquad \text{for } j = 1,\ldots, n \tag{11.51}$$

so that the term T_k in (11.49) becomes the following function τ_k of $\mathbf{y}$:

$$\tau_k = \alpha_k \prod_{j=1}^{n} (e^{y_j})^{a_{kj}} = \alpha_k e^{\mathbf{a}_k^t \mathbf{y}} \qquad \text{for } k = 1,\ldots, M, \tag{11.52}$$

where $\mathbf{a}_k = (a_{k1},\ldots,a_{kn})^t$ for $k = 1,\ldots, M$. Furthermore, in addition to using the substitution (11.51) in Problem GP, let us also equivalently write the objective function of GP as one of minimizing $\ln[f(\mathbf{x})]$ and the constraints of GP as $\ln[g_i(\mathbf{x})] \le 0$ for $i = 1,\ldots, m$, noting the monotonicity of the logarithmic function and the positivity of the objective and constraint functions. Hence, applying the transformation (11.51) to this representation of GP, we equivalently derive the following problem:

$$\text{Minimize } \ln[F(\mathbf{y})] \tag{11.53a}$$

$$\text{subject to } \ln[G_i(\mathbf{y})] \le 0 \qquad \text{for } i = 1,\ldots, m \tag{11.53b}$$

$$\mathbf{y} \text{ unrestricted in sign,}$$

where from (11.50)–(11.52) we have

$$F(\mathbf{y}) \equiv \sum_{k \in J_0} \tau_k \quad \text{and} \quad G_i(\mathbf{y}) \equiv \sum_{k \in J_i} \tau_k \qquad \text{for } i = 1,\ldots, m. \tag{11.53c}$$

The following result establishes an extremely useful characterization of Problem (11.53).

11.5.1 Lemma

Given the posynomial geometric programming problem GP, consider the equivalent problem (11.53) obtained under the transformation (11.51). Then the objective and constraint functions of this problem are all convex, and hence, (11.53) is a convex programming problem.

Proof

First, consider any term τ_k, which is a function of y, as defined by (11.52). Denoting $\mathbf{a}_k = (a_{k1},\dots,a_{kn})^t$, we have $\nabla\tau_k = \tau_k\mathbf{a}_k$ and $\nabla^2\tau_k = \tau_k\mathbf{a}_k\mathbf{a}_k^t$, where ∇ and ∇^2 denote the gradient and Hessian operators, respectively. Now denote $h(\mathbf{y}) = \ln[F(\mathbf{y})]$. We have $\nabla h(\mathbf{y}) = \nabla F(\mathbf{y})/F(\mathbf{y})$ and

$$\nabla^2 h(\mathbf{y}) = \frac{[F(\mathbf{y})\nabla^2 F(\mathbf{y}) - \nabla F(\mathbf{y})\nabla F(\mathbf{y})^t]}{[F(\mathbf{y})]^2}$$

$$= \frac{\left[\sum_{k\in J_0}\tau_k\right]\left[\sum_{k\in J_0}\tau_k\mathbf{a}_k\mathbf{a}_k^t\right] - \left[\sum_{k\in J_0}\tau_k\mathbf{a}_k\right]\left[\sum_{k\in J_0}\tau_k\mathbf{a}_k^t\right]}{[F(\mathbf{y})]^2}.$$

Using (11.53c) and the foregoing expressions for $\nabla\tau_k$ and $\nabla^2\tau_k$, the numerator of $\nabla^2 h(\mathbf{y})$ equals

$$\sum_{k\in J_0}\sum_{\ell\in J_0}\tau_k\tau_\ell\mathbf{a}_k\mathbf{a}_k^t - \sum_{k\in J_0}\sum_{\ell\in J_0}\tau_k\tau_\ell\mathbf{a}_k\mathbf{a}_\ell^t = \sum\sum_{k<\ell \text{ in } J_0}\tau_k\tau_\ell[\mathbf{a}_k\mathbf{a}_k^t + \mathbf{a}_\ell\mathbf{a}_\ell^t]$$

$$- \sum\sum_{k<\ell \text{ in } J_0}\tau_k\tau_\ell[\mathbf{a}_k\mathbf{a}_\ell^t + \mathbf{a}_\ell\mathbf{a}_k^t] = \sum\sum_{k<\ell \text{ in } J_0}\tau_k\tau_\ell(\mathbf{a}_k - \mathbf{a}_\ell)(\mathbf{a}_k - \mathbf{a}_\ell)^t.$$

Consequently, as $\tau_k > 0$ and $\tau_\ell > 0$, and $(\mathbf{a}_k - \mathbf{a}_\ell)(\mathbf{a}_k - \mathbf{a}_\ell)^t$ is positive semidefinite, we have that $\ln[F(\mathbf{y})]$ is a convex function. Similarly, $\ln[G_i(\mathbf{y})]$ is a convex function for each $i = 1,\dots, m$, and this completes the proof.

Now, assuming that a suitable constraint qualification holds true (such as the interiority constraint qualification of Theorem 6.2.4), we can invoke Theorem 6.2.4 to assert that there is no duality gap between (11.53) and its Lagrangian dual stated below:

$$\text{LD:} \quad \text{Maximize} \quad \{\theta(\mathbf{u}) : \mathbf{u} \ge 0\}, \tag{11.54a}$$

where

$$\theta(\mathbf{u}) = \min_{\mathbf{y}}\{L(\mathbf{y},\mathbf{u})\} \tag{11.54b}$$

and where $L(\mathbf{y}, \mathbf{u})$ is the Lagrangian function given by

$$L(\mathbf{y},\mathbf{u})=\ln[F(\mathbf{y})]+\sum_{i=1}^{m} u_i \ln[G_i(\mathbf{y})]. \tag{11.54c}$$

Since for any u ≥ 0, $\theta(\mathbf{u})$ equals that value of $L(\mathbf{y}, \mathbf{u})$ evaluated at the point y for which $\nabla_{\mathbf{y}} L(\mathbf{y},\mathbf{u})$ equals zero by Lemma 11.5.1, we can equivalently write the Lagrangian dual (11.54) as follows:

$$\text{Maximize } \{L(\mathbf{y}, \mathbf{u}) : \nabla_{\mathbf{y}} L(\mathbf{y},\mathbf{u}) = \mathbf{0},\ \mathbf{u} \geq \mathbf{0},\ \mathbf{y} \text{ unrestricted}\}. \tag{11.55}$$

To simplify (11.55) further, note that as in the proof of Lemma 11.5.1, we have

$$\begin{aligned}\nabla_{\mathbf{y}} L(\mathbf{y},\mathbf{u}) &= \frac{\nabla F(\mathbf{y})}{F(\mathbf{y})}+\sum_{i=1}^{m} u_i \frac{\nabla G_i(\mathbf{y})}{G_i(\mathbf{y})} \\ &= \frac{1}{F(\mathbf{y})}\sum_{k\in J_0} \tau_k \mathbf{a}_k + \sum_{i=1}^{m}\frac{u_i}{G_i(\mathbf{y})}\left(\sum_{k\in J_i} \tau_k \mathbf{a}_k\right).\end{aligned} \tag{11.56}$$

We now employ a second transformation. Define $\delta_1,\ldots,\delta_M$ according to

$$\delta_k = \frac{\tau_k}{F} \text{ for all } k \in J_0 \qquad \text{and} \qquad \delta_k = \frac{u_i \tau_k}{G_i} \text{ for all } k \in J_i,\ i = 1,\ldots, m. \tag{11.57}$$

Note that we have dropped the argument (y) for notational convenience, recognizing the dependence of F, G_i, τ_k, and δ_k for all $i = 1,\ldots, m$, $k = 1,\ldots, M$ on y. However, we would now like to treat $\delta = (\delta_1,\ldots,\delta_M)^t$ as a set of *variables* and write (11.55) in terms of (δ, u) by eliminating y from the problem.

Note that, from (11.53c) and (11.57), we must have

$$\sum_{k\in J_0} \delta_k = 1 \qquad \text{and} \qquad \sum_{k\in J_i} \delta_k = u_i \qquad \text{for } i = 1,\ldots, m. \tag{11.58}$$

Constraints (11.58) are called *normalization constraints* and, together with $\boldsymbol{\delta} \geq \mathbf{0}$, restrict the values that δ can assume in (11.57). Furthermore, using (11.56), the equality constraint in (11.55) can be written as follows under the transformation (11.57):

$$\sum_{k=1}^{M} \delta_k \mathbf{a}_k = \mathbf{0}. \tag{11.59}$$

Constraint (11.59) is known as the *orthogonality constraint*, since it asserts that δ is orthogonal to each of the n rows of the $n \times M$ matrix A having columns $\mathbf{a}_1,\ldots,\mathbf{a}_M$.

In the transformed problem, we impose the relationships (11.58) and (11.59) on (δ, u), along with nonnegativity restrictions. However, given any (δ,

u) feasible to these conditions, since there does not necessarily exist a y that satisfies (11.57), we need some further analysis to justify the utility of the transformed problem derived below.

Toward this end, let us also simplify the objective function of (11.55) under (11.57) and (11.58). Consider the term $u_i \ln[G_i]$ for any $i \in \{1,..., m\}$. Assuming that $u_i > 0$, and writing this term as $u_i \ln(u_i)+u_i \ln[G_i/u_i]$, we get, upon using (11.58), (11.57), and (11.52) in turn, that

$$u_i \ln[G_i] = u_i \ln(u_i) + u_i \ln\left[\frac{G_i}{u_i}\right] = u_i \ln(u_i) + \sum_{k\in J_i} \delta_k \ln\left[\frac{G_i}{u_i}\right]$$

$$= u_i \ln(u_i) + \sum_{k\in J_i} \delta_k \ln\left[\frac{\tau_k}{\delta_k}\right] = u_i \ln(u_i) + \sum_{k\in J_i} \delta_k \ln\left(\frac{\alpha_k}{\delta_k} e^{\mathbf{a}_k^t \mathbf{y}}\right) \quad (11.60a)$$

$$= u_i \ln(u_i) + \sum_{k\in J_i} \delta_k \ln\left[\frac{\alpha_k}{\delta_k}\right] + \sum_{k\in J_i} \delta_k \mathbf{a}_k^t \mathbf{y}.$$

Similarly, we have

$$\ln[F] = \sum_{k\in J_0} \delta_k \ln\left[\frac{\alpha_k}{\delta_k}\right] + \sum_{k\in J_0} \delta_k \mathbf{a}_k^t \mathbf{y}. \quad (11.60b)$$

Hence, noting (11.50b) and that $\sum_{k=1}^{M} \delta_k \mathbf{a}_k^t \mathbf{y} = 0$ by (11.59), we observe from (11.54c) and (11.60) that the objective function in (11.55) is given by

$$\sum_{k=1}^{M} \delta_k \ln\left[\frac{\alpha_k}{\delta_k}\right] + \sum_{i=1}^{m} u_i \ln(u_i). \quad (11.61)$$

Noting that $u_i \ln(u_i) \to 0$ as $u_i \to 0^+$, and that $\delta_k \ln[\alpha_k/\delta_k] \to 0$ as $\delta_k \to 0^+$, we finally use (11.58), (11.59), and (11.61) to replace (11.55) with the following *dual geometric program* (DGP) in the variables (δ, **u**), where the separable objective function terms are defined as zeros for $\delta_k = 0$ or $u_i = 0$:

$$\text{DGP:} \quad \text{Maximize} \quad \sum_{k=1}^{M} \delta_k \ln\left[\frac{\alpha_k}{\delta_k}\right] + \sum_{i=1}^{m} u_i \ln(u_i) \quad (11.62a)$$

$$\text{subject to} \quad \sum_{k=1}^{M} \delta_k \mathbf{a}_k \equiv \mathbf{A}\delta = \mathbf{0} \quad (11.62b)$$

$$\sum_{k\in J_0} \delta_k = 1 \quad (11.62c)$$

$$\sum_{k\in J_i} \delta_k = u_i \qquad \text{for } i = 1,..., m \quad (11.62d)$$

$$\delta_k \geq 0 \qquad \text{for } k = 1,..., M \quad (11.62e)$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots, m. \qquad (11.62f)$$

Note that Problem DGP is a linearly constrained problem having a separable, concave objective function (see Exercise 11.44) and is therefore a convex programming problem that is readily solvable by using the methods of Chapter 10 and Section 11.3. Note from (11.62d) that we can write the variables u_i, $i = 1,\ldots, m$, in terms of the δ-variables. Furthermore, assuming that the $(n + 1)$ constraints (11.62b) and (11.62c) are linearly independent, we can solve for some $(n + 1)$ variables δ_k in terms of the remaining $(M - n - 1)$ δ-variables. The resulting degrees of freedom in the problem due to (11.62b)–(11.62d) is called the

$$\textit{degree of difficulty}\ (\text{DD}) \equiv \text{number of terms } (M) - \text{number of variables } (n) - 1. \qquad (11.63)$$

In general, the degree of difficulty DD equals M minus the number of linearly independent constraints in (11.62b, c). Note that if DD = 0, as is sometimes the case, then the solution to DGP, if it exists, is determined uniquely by (11.62b)–(11.62d) itself. Otherwise, a linearly constrained problem that is essentially embedded in dimension DD needs to be solved. The following result prescribes the recovery of the optimum to GP from an optimum to DGP.

11.5.2 Theorem

Consider the dual geometric program DGP, and suppose that $(\delta^*, \mathbf{u}^*) > 0$ solves this problem with an optimal objective function value v(DGP). Furthermore, let $\mathbf{v}^* = (v_1^*,\ldots,v_n^*)^t$, and let $\mathbf{w}^* = (w_0^*, w_1^*,\ldots, w_m^*)^t$ be the corresponding optimal Lagrange multiplier values associated with the constraints (11.62b) and (11.62c, d), respectively. Then an optimum $\mathbf{y}^*$ to Problem (11.53) is given by

$$y_j^* = v_j^* \qquad \text{for } j = 1,\ldots, n, \qquad (11.64a)$$

with the optimal objective value of this problem being v(DGP) and with $\mathbf{u}^*$ being the set of optimal Lagrange multipliers associated with (11.53b). Moreover,

$$x_j^* = e^{y_j^*} \qquad \text{for } j = 1,\ldots, n \qquad (11.64b)$$

solves Problem GP.

Proof

Since $(\delta^*, u^*) > 0$ at optimality, by the differentiability of the objective function at this point and the linearity of the constraints, Lemma 5.1.4 asserts that we must have a solution to the KKT system given by (11.62b)–(11.62f), along with the complementary slack dual feasibility conditions

$$\left(\ln\left[\frac{\alpha_k}{\delta_k^*}\right]-1\right)+\mathbf{a}_k^t\mathbf{v}^*+w_i^*=0 \qquad \text{for all } k\in J_i,\ i=0,1,\ldots,m \qquad (11.65)$$

$$[\ln(u_i^*)+1]-w_i^*=0 \qquad \text{for all } i=1,\ldots,m, \qquad (11.66)$$

where $\mathbf{v}^*$ and w^* are as defined in the theorem. Substituting for w_i^* from (11.66) into (11.65) for $i = 1,\ldots, m$, we get

$$\mathbf{a}_k^t\mathbf{v}^*=\ln\left[\frac{\delta_k^*}{\alpha_k u_i^*}\right] \qquad \text{for all } k\in J_i,\ i=1,\ldots,m. \qquad (11.67)$$

Now consider y^* as given by (11.64a). We then have from (11.52), (11.53c), (11.62d), and (11.67) that

$$\begin{aligned} G_i(y^*) &= \sum_{k\in J_i}\tau_k=\sum_{k\in J_i}\alpha_k e^{\mathbf{a}_k^t y^*} \\ &= \sum_{k\in J_i}\alpha_k\frac{\delta_k^*}{\alpha_k u_i^*}=1 \qquad \text{for } i=1,\ldots,m. \end{aligned} \qquad (11.68)$$

Moreover, from (11.52) and (11.65)–(11.67), we have

$$\tau_k=\alpha_k e^{\mathbf{a}_k^t y^*}=\frac{\delta_k^*}{u_i^*} \quad \text{for all } k\in J_i,\ i=1,\ldots,m. \qquad (11.69a)$$

Furthermore, we have

$$\tau_k=\alpha_k e^{\mathbf{a}_k^t y^*}=\alpha_k e^{[1-w_0^*+\ln(\delta_k^*/\alpha_k)]}=\delta_k^* e^{(1-w_0^*)} \qquad \text{for all } k\in J_0.$$

But $F(y^*)=\sum_{k\in J_0}\tau_k=e^{(1-w_0^*)}\sum_{k\in J_0}\delta_k^*=e^{1-w_0^*}$ from (11.62c). Hence, we have

$$\tau_k=F(y^*)\delta_k^* \qquad \text{for all } k\in J_0. \qquad (11.69b)$$

Substituting (11.69) into (11.56) and using (11.62b) and (11.68), we get

$$\nabla_y L(\mathbf{y}^*, \mathbf{u}^*) = \sum_{k \in J_0} \delta_k^* \mathbf{a}_k - \sum_{i=1}^{m} \sum_{k \in J_i} \delta_k^* \mathbf{a}_k = 0. \tag{11.70}$$

Consequently, from (11.62e, f), (11.68), and (11.70) the primal-dual solution $(\mathbf{y}^*, \mathbf{u}^*)$ satisfies the KKT conditions for Problem (11.53) and, hence, using Lemma 11.5.1, solves this problem. Moreover, noting (11.54c), (11.60), (11.61), (11.68), and (11.69), we have

$$v(\text{DGP}) = L(\mathbf{y}^*, \mathbf{u}^*) = \ln[F(\mathbf{y}^*)].$$

Finally, by the equivalence of GP and Problem (11.53) under the transformation (11.51), we also have that $x_j^* = e^{y_j^*}$, $j = 1,\ldots,n$, solves GP, and this completes the proof.

Observe that given a positive optimal solution to Problem DGP, we are able to claim that GP has an optimum and, moreover, recover an optimum to this problem via (11.64). On the other hand, if GP has an optimum, and if the interior point constraint qualification of Theorem 6.2.4 holds true, then it can be shown that DGP also has an optimum $(\delta^*, \mathbf{u}^*)$ with the same objective function value v^*, and that an optimum to Problem (11.53) can be recovered by solving the system

$$\mathbf{a}_k^t \mathbf{y} = \ln\left[\frac{\delta_k^* e^{v^*}}{\alpha_k}\right] \qquad \text{for } k \in J_0 \tag{11.71a}$$

$$\mathbf{a}_k^t \mathbf{y} = \ln\left[\frac{\delta_k^*}{u_i^* \alpha_k}\right] \qquad \text{for } k \in J_i \text{ and } i \in \{1,\ldots, m\} \text{ such that } u_i^* > 0. \tag{11.71b}$$

This system arises from (11.52) and (11.57), noting that $\ln[F(\mathbf{y}^*)] = v^*$ and that $G_i(\mathbf{y}^*) = 1$ for the active constraints that have $u_i^* > 0$ (see Duffin et al. [1967]). Note from (11.52) and (11.69) that the proof of Theorem 11.52 verifies this system to yield $\mathbf{y}^*$ in terms of the (primal) solution to DGP under the conditions of Theorem 11.5.2. Hence, (11.71) provides an alternative to (11.64a) for recovering a primal optimal solution to Problem GP via (11.64b).

11.5.3 Example (Zero Degrees of Difficulty)

Suppose that we wish to construct a right circular cylinder of radius r and height h that is closed at both ends, has a volume of at least V, and uses the least amount of material. Hence, the problem we wish to solve is to minimize the total surface area $2\pi r^2 + 2\pi rh$, so that the volume $\pi r^2 h$ is at least V. Rewriting the constraint in standard form, this gives

$$\text{GP: Minimize}\left\{2\pi r^2 + 2\pi rh : \frac{V}{\pi} r^{-2}h^{-1} \le 1, r > 0, h > 0\right\}.$$

Note that the number of terms in the problem is $M = 3$ and that the number of variables is $n = 2$, namely, r and h. Hence, from (11.63) the degree of difficulty equals zero. The α coefficients for the three terms are given by $\alpha_1 = 2\pi$, $\alpha_2 = 2\pi$, and $\alpha_3 = V/\pi$. The respective exponent vectors are

$$\mathbf{a}_1^t = (2, 0), \qquad \mathbf{a}_2^t = (1, 1), \qquad \text{and} \qquad \mathbf{a}_3^t = (-2, -1).$$

The corresponding orthogonality and normalization constraints (11.62b, c, d) are as follows, noting that $J_0 = \{1, 2\}$, and $J_1 = \{3\}$:

$$\begin{aligned} 2\delta_1 + \delta_2 - 2\delta_3 &= 0 \\ \delta_2 - \delta_3 &= 0 \\ \delta_1 + \delta_2 \qquad &= 1 \\ \delta_3 &= u_1. \end{aligned}$$

Solving, we obtain $\delta_1^* = 1/3$, $\delta_2^* = \delta_3^* = u_1^* = 2/3$. Note that $(\delta^*, \mathbf{u}^*) > \mathbf{0}$, so that the condition of Theorem 11.5.2 holds true. The optimal objective function value of Problem DGP is $v^* = (1/3)\ln[6\pi] + (2/3)\ln[3\pi] + (2/3)\ln(3V/2\pi) + (2/3)\ln(2/3) = \ln\left[(54\pi V^2)^{1/3}\right]$. Hence, $e^{v^*} = (54\pi V^2)^{1/3}$. Consequently, from (11.71), we get

$$2y_1 = \ln\left[\frac{1}{3}\frac{(54\pi V^2)^{1/3}}{2\pi}\right] = \ln\left[\left(\frac{V}{2\pi}\right)^{2/3}\right]$$

$$y_1 + y_2 = \ln\left[\frac{2}{3}\frac{(54\pi V^2)^{1/3}}{2\pi}\right] = \ln\left[\left(\frac{2V^2}{\pi^2}\right)^{1/3}\right]$$

$$-2y_1 - y_2 = \ln\left[\frac{\pi}{V}\right].$$

(Note that the third equation above is redundant.) Solving, we get $y_1 = \ln\left[(V/2\pi)^{1/3}\right]$ and $y_2 = \ln\left[(4V/\pi)^{1/3}\right]$. Hence, from (11.64b), we get $r^* = (V/2\pi)^{1/3}$ and $h^* = (4V/\pi)^{1/3} = 2r^*$ as the optimum for Problem GP.

11.5.4 Example (Degree of Difficulty = 1)

Consider Example 11.5.3, and suppose now that we also need to connect a wire of length h joining the centers of the base and the top of the cylinder. The ratio of the cost per unit length (cm) of this wire to the cost per unit surface area (cm^2) of the cylinder is 2π. Also, the volume is required to be at least V = $(256\pi/135)$ cm^3.

Problem GP now has the form:

$$\text{Minimize}\left\{2\pi r^2+2\pi rh+2\pi h:\frac{V}{\pi}r^{-2}h^{-1}\le 1,\, r>0,\, h>0\right\}.$$

Here, we now have $m = 1$, $n = 2$, $M = 4$, DD $= M - n - 1 = 1$, $\alpha_1 = 2\pi$, $\alpha_2 = 2\pi$, $\alpha_3 = 2\pi$, $\alpha_4 = V/\pi$, $\mathbf{a}_1^t = (2, 0)$, $\mathbf{a}_2^t = (1, 1)$, $\mathbf{a}_3^t = (0, 1)$, $\mathbf{a}_4^t = (-2, -1)$, with J_0 = $\{1, 2, 3\}$ and $J_1 = \{4\}$. The orthogonality and normalization constraints (11.62b)–(11.62d) give

$$2\delta_1+\delta_2-2\delta_4=0,\ \delta_2+\delta_3-\delta_4=0,\ \delta_1+\delta_2+\delta_3=1,\ \delta_4=u_1.$$

Solving for all variables in terms of δ_4, which represents the single degree of freedom, we obtain

$$\delta_1=(1-\delta_4),\ \delta_2=(4\delta_4-2),\ \delta_3=(2-3\delta_4),\text{ and } u_1=\delta_4. \qquad (11.72)$$

The nonnegativity constraints (11.62e, f) then imply that $1/2 \le \delta_4 \le 2/3$. Hence, Problem DGP, projected onto the space of the variable δ_4, is given as follows, where we have used $\delta_4 \ln(\alpha_4/\delta_4) + u_1 \ln(u_1) = \delta_4 \ln(\alpha_4)$, since $u_1 = \delta_4$ in (11.72):

$$\begin{aligned}\text{Maximize}\quad &(1-\delta_4)\ln\left[\frac{2\pi}{1-\delta_4}\right]+(4\delta_4-2)\ln\left[\frac{2\pi}{4\delta_4-2}\right]\\ &+(2-3\delta_4)\ln\left[\frac{2\pi}{2-3\delta_4}\right]+\delta_4\ln\left[\frac{256}{135}\right]\\ \text{subject to}\quad &\frac{1}{2}\le\delta_4\le\frac{2}{3}.\end{aligned}$$

[Note that we could have opted to solve (11.62) directly, without projecting it first into a one-dimensional problem.] Now, differentiating the objective function of DGP and setting it equal to zero gives $\delta_4 = 7/12$. Since the objective function is concave and this value is feasible, it solves Problem DGP. Using (11.72), we obtain

$$\delta_1^*=\frac{5}{12},\quad \delta_2^*=\frac{1}{3},\quad \delta_3^*=\frac{1}{4},\quad \delta_4^*=\frac{7}{12},\quad\text{and}\quad u_1^*=\frac{7}{12},$$

which satisfy the condition of Theorem 11.5.2. The optimal objective function value is $v(\text{DGP}) \equiv v^* = \ln[8.53333\pi]$. Hence, $e^{v^*} = 8.53333\pi$. Consequently, from (11.71) we obtain $2y_1^* = \ln[(5/12)(8.53333\pi)(1/2\pi)]$ using $k = 1$, and we get $y_2^* = \ln[(1/4)(8.53333\pi)(1/2\pi)]$ using $k = 3$. [The other equations in (11.71) are redundant.] Using (11.64b), this finally yields

$$r^* = e^{y_1^*} = 1.33333 \text{ cm}, \qquad \text{and} \qquad h^* = e^{y_2^*} = 1.06667 \text{ cm}.$$

In Exercise 11.45 we ask the reader to study the sensitivity of the solution to the cost ratio factor specified in the objective function of Problem GP.

Exercises

[11.1] Consider the following linear programming problem:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^t\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

a. Write the KKT system for this problem.
b. Use the complementary pivoting algorithm to solve the KKT system for the following problem:

$$\begin{aligned} \text{Minimize} \quad & -x_1 - 3x_2 \\ \text{subject to} \quad & 2x_1 + 3x_2 \leq 6 \\ & -x_1 + 2x_2 \leq 2 \\ & x_1, x_2 \geq 0. \end{aligned}$$

c. Repeat Part b if the first constraint is replaced by $x_2 \leq 2$.

[11.2] Consider the linear complementary problem to find (**w**, **z**) such that $\mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q}$, $\mathbf{w}^t\mathbf{z} = 0$, and $\mathbf{w}, \mathbf{z} \geq \mathbf{0}$, where

$$\mathbf{M} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 1 & -1 & -2 \\ 1 & 0 & 1 & -2 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}.$$

a. Is the matrix **M** copositive plus?
b. Apply Lemke's algorithm discussed in Section 11.1 to the above problem.

[11.3] Find a complementary basic feasible solution to the system $\mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q}$, $\mathbf{w}^t\mathbf{z} = 0$, and $\mathbf{w}, \mathbf{z} \geq \mathbf{0}$ by using Lemke's algorithm. Here

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 2 & 3 & 2 & 2 \\ 1 & 4 & 1 & 4 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 1 \\ -8 \\ -8 \\ 2 \end{bmatrix}.$$

[11.4] Consider the LCP problem of finding a solution, if one exists, to the system $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $\mathbf{w} \geq \mathbf{0}$, $\mathbf{z} \geq \mathbf{0}$, and $\mathbf{w}^t\mathbf{z} = 0$, where $\mathbf{M}$ is a $p \times p$ matrix. Define

$$h(\mathbf{y}) = \min\left\{ \sum_{j=1}^{n} y_j w_j + (1-y_j)z_j : \mathbf{w} - \mathbf{Mz} = \mathbf{q},\ \mathbf{w} \geq \mathbf{0}, \text{ and } \mathbf{z} \geq \mathbf{0} \right\}. \quad (11.73)$$

a. Show that h is a concave function of $\mathbf{y}$ over $\mathbf{0} \leq \mathbf{y} \leq \mathbf{e}$, where $\mathbf{e}$ is a vector of p ones.
b. Show that LCP is equivalent to minimizing h over either $\mathbf{0} \leq \mathbf{y} \leq \mathbf{e}$ or $\mathbf{y}$ restricted to take on binary values.
c. Assume that the set $Z = \{\mathbf{z} : -\mathbf{Mz} \leq \mathbf{q},\ \mathbf{z} \geq \mathbf{0}\}$ is nonempty and bounded, with $0 \leq z_k^+ \equiv \max\{z_k : \mathbf{z} \in Z\} < \infty$ for all $k = 1,\ldots,p$. Let $\mathbf{M}_j$ denote the jth row of $\mathbf{M}$, $j = 1,\ldots,p$. Construct the set Z_p by multiplying each of the inequalities in Z with each y_j and $1 - y_j$ for $j = 1,\ldots,p$. Hence, construct the problem

$$\text{LCP}': \text{ Minimize}\left\{ \sum_{j=1}^{n} y_j(q_j + \mathbf{M}_j\mathbf{z}) + \sum_{j=1}^{n} (1-y_j)z_j : (\mathbf{y},\mathbf{z}) \in Z_p, \mathbf{y} \text{ binary} \right\}.$$

Now linearize LCP′ by substituting x_{ij} in place of the product $y_i z_j$ for all $i, j = 1,\ldots,p$, and hence obtain a resulting linear mixed-integer zero-one programming problem MIP in continuous variables $\mathbf{z}$ and $\mathbf{x}$ and in binary variables $\mathbf{y}$. Show that the constraints of MIP imply that

$$0 \leq x_{ij} \leq z_j^+ y_i \quad \text{and} \quad z_j - z_j^+(1-y_i) \leq x_{ij} \leq z_j \quad \text{for all } i, j = 1,\ldots,p.$$

Hence show that solving MIP is equivalent to solving LCP.
d. Discuss how you might use Parts b and c to derive a solution method for solving LCP. (Sherali et al. [1991a,b] discuss this transformation and related algorithms.)

[11.5] Consider the problem to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{Hx}$ subject to $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, where $\mathbf{A}$ is $m \times n$ and where $\mathbf{H}$ is $n \times n$ and symmetric. Now, consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^t\mathbf{x} - \mathbf{b}^t\mathbf{u} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} - \mathbf{v} = -\mathbf{c} \\ & \mathbf{v}^t\mathbf{x} = 0 \\ & \mathbf{x}, \mathbf{v} \geq \mathbf{0}, \qquad \mathbf{u} \text{ unrestricted.} \end{aligned}$$

a. Show that an optimal solution to the problem stated above gives a point with minimal objective value among all the KKT points. Does this imply that the optimal solution of the problem is a global minimum?
b. Give an interpretation of the objective function of the above problem.
c. Suggest a procedure for solving the above problem using the technique of Exercise 11.4, for example, and illustrate by solving the following problem:

$$\begin{aligned} \text{Minimize} \quad & -(x_1-2)^2 - (x_2-2)^2 \\ \text{subject to} \quad & -2x_1 + x_2 + x_3 = 4 \\ & 3x_1 + 2x_2 + x_4 = 12 \\ & 3x_1 - 2x_2 + x_5 = 6 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0. \end{aligned}$$

[11.6] Consider the LCP problem of finding a solution, if one exists, to the system $\mathbf{Mz} + \mathbf{q} \geq \mathbf{0}$, $\mathbf{z} \geq \mathbf{0}$, and $(\mathbf{Mz} + \mathbf{q})^t\mathbf{z} = 0$, where $\mathbf{M}$ is a $p \times p$ matrix. Consider the following linear mixed-integer programming problem, where e is a vector of p ones:

$$\begin{aligned} \text{MIP: Minimize} \quad & \alpha \\ \text{subject to} \quad & \mathbf{0} \leq \mathbf{Mx} + \alpha\mathbf{q} \leq \mathbf{e} - \mathbf{y} \\ & \mathbf{0} \leq \mathbf{x} \leq \mathbf{y} \\ & \mathbf{y} \text{ binary}, 0 \leq \alpha \leq 1. \end{aligned}$$

Show that if MIP has an optimum solution $(\alpha^*, \mathbf{x}^*, \mathbf{y}^*)$ with objective value $\alpha^* > 0$, then $\mathbf{z} = \mathbf{x}/\alpha^*$ solves LCP. On the other hand, if $\alpha^* = 0$ at optimality, then show that LCP has no solution. (This formulation is due to Pardalos and Rosen [1988].)

[11.7] Use the complementary pivoting algorithm to solve the following quadratic programming problem:

$$\begin{aligned} \text{Maximize} \quad & 4x_1 - 2x_2 - 3x_1^2 - 3x_1x_2 - 2x_2^2 \\ \text{subject to} \quad & 3x_1 + 2x_2 \le 6 \\ & -x_1 + 2x_2 \le 4 \\ & x_1, x_2 \ge 0. \end{aligned}$$

[11.8] Solve the KKT system for the following problem by the complementary pivoting algorithm:

$$\begin{aligned} \text{Minimize} \quad & -3x_1 + 2x_2 - 4x_3 + 3x_1^2 + 2x_2^2 + 6x_3^2 - x_1x_2 - 2x_1x_3 + 3x_1x_3 \\ \text{subject to} \quad & 2x_1 + x_2 + x_3 \ge 4 \\ & x_1 + 2x_2 + x_3 \le 8 \\ & -3x_1 + 2x_2 \le -4 \\ & x_1, x_2, x_3 \ge 0. \end{aligned}$$

[11.9] Consider the LCP problem of finding a solution, if one exists, to the system $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $\mathbf{w} \ge \mathbf{0}$, $\mathbf{z} \ge \mathbf{0}$, and $\mathbf{w}^t\mathbf{z} = 0$, where $\mathbf{M}$ is a $p \times p$ matrix. Define $Z = \{\mathbf{z} \ge \mathbf{0} : \mathbf{Mz} + \mathbf{q} \ge \mathbf{0}\}$, and $W = \{\mathbf{w} : \mathbf{0} \le \mathbf{w} \le K\mathbf{e}\}$, where K is a large number and $\mathbf{e}$ is a vector of p ones. Consider the problem

$$\text{LCP}': \quad \text{Minimize}\left\{ \sum_{j=1}^{n} [\min\{0, w_j\} + z_j] : \mathbf{z} \in Z, \mathbf{w} \in W, \mathbf{w} + \mathbf{z} = \mathbf{q} + \mathbf{Mz} \right\}.$$

Discuss the structure of LCP′ and its equivalence with respect to solving LCP. (This formulation is proposed by Bard and Falk [1982].)

[11.10] In Section 11.1 we showed constructively in Theorem 11.1.8 that if the system $\mathbf{w} - \mathbf{Mz} = \mathbf{q}$, $(\mathbf{w}, \mathbf{z}) \ge \mathbf{0}$ is consistent and if $\mathbf{M}$ is copositive plus, then the system defined by (11.1), (11.2), and (11.3) is solvable. Prove this fact directly.

[11.11] In this exercise we describe the *principal pivoting method* credited to Cottle and Dantzig [1968] for solving the following linear complementary problem:

$$\begin{aligned} \mathbf{w} - \mathbf{Mz} &= \mathbf{q} \\ \mathbf{w}, \mathbf{z} &\ge \mathbf{0} \\ \mathbf{w}^t\mathbf{z} &= 0. \end{aligned}$$

If the system has a solution, if $\mathbf{M}$ is positive definite, and if every basic solution to the above system is nondegenerate, then the algorithm stops in a finite number of steps with a complementary basic feasible solution.

Initialization Step Consider the basic solution $\mathbf{w} = \mathbf{q}$, $\mathbf{z} = \mathbf{0}$, and construct the associated tableau. Go to the Main Step.

Main Step

1. Let (w, z) be a complementary basic solution with $\mathbf{z} \geq \mathbf{0}$. If $\mathbf{w} \geq \mathbf{0}$, stop; (w, z) is a complementary basic feasible solution. Otherwise, let $w_k < 0$. Let v be the variable complementary to w_k and go to Step 2.
2. Increase v until either w_k reaches value zero or some positive basic variable decreases to zero. In the former case, go to Step 1 after pivoting to update the tableau. In the latter case, pivot to update the tableau, and let v be the variable complementary to that just removed from the basis. Repeat Step 2.

a. Show that at each iteration of Step 2, w_k increases until it reaches the value zero.
b. Prove finite convergence of the algorithm to a complementary basic feasible solution.
c. Can the method be used to solve a quadratic program where the objective function is strictly convex?

[11.12] In a *bimatrix game*, there are two players, I and II. For Player I there exist m possible strategies, and for Player II there exist n possible strategies. If Player I chooses strategy i and Player II chooses strategy j, then Player I loses a_{ij} and Player II loses b_{ij}. Let the loss matrices of Players I and II be **A** and **B**, where a_{ij} and b_{ij} are the (i, j)th entries of **A** and **B**, respectively. If Player I chooses to play strategy i with probability x_i and Player II chooses to play strategy j with probability y_j, then the expected losses of the two players are $\mathbf{x}^t\mathbf{A}\mathbf{y}$ and $\mathbf{x}^t\mathbf{B}\mathbf{y}$, respectively. The strategy pair $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is said to be an equilibrium point if

$$\bar{\mathbf{x}}^t\mathbf{A}\bar{\mathbf{y}} \leq \mathbf{x}^t\mathbf{A}\bar{\mathbf{y}} \qquad \text{for all } \mathbf{x} \geq \mathbf{0} \text{ such that } \sum_{i=1}^{m} x_i = 1$$

$$\bar{\mathbf{x}}^t\mathbf{B}\bar{\mathbf{y}} \leq \bar{\mathbf{x}}^t\mathbf{B}\mathbf{y} \qquad \text{for all } \mathbf{y} \geq \mathbf{0} \text{ such that } \sum_{j=1}^{n} y_j = 1.$$

a. Show how an equilibrium pair $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is obtained by formulating a suitable linear complementary problem of the form $\mathbf{w} - \mathbf{M}\mathbf{z} = \mathbf{q}$, $\mathbf{w}^t\mathbf{z} = 0$, and $\mathbf{w}, \mathbf{z} \geq \mathbf{0}$.
b. Investigate the properties of the matrix **M**. Verify whether the complementary problem has a solution.
c. Find an equilibrium pair for the following loss matrices:

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 3 \\ 1 & 3 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 4 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$

[11.13] The following problem is usually referred to as the *nonlinear complementary problem*. Find a point $\mathbf{x} \in R^n$ such that $\mathbf{x} \geq \mathbf{0}$, $\mathbf{g}(\mathbf{x}) \geq \mathbf{0}$, and $\mathbf{x}^t\mathbf{g}(\mathbf{x}) = 0$, where $\mathbf{g}: R^n \to R^n$ is a continuous vector function.

a. Show that the linear complementary problem is a special case of the above nonlinear problem.
b. Show that the KKT conditions for optimality for a nonlinear programming problem could be written as a nonlinear complementary problem.
c. Show that if $\mathbf{g}$ satisfies the following strong monotonicity property, then there exists a unique solution to the nonlinear complementary problem. (Detailed proof is given in Karamardian [1969].) We say that $\mathbf{g}$ is *strongly monotone* if there exists an $\varepsilon > 0$ such that

$$(\mathbf{y}-\mathbf{x})^t[\mathbf{g}(\mathbf{y})-\mathbf{g}(\mathbf{x})] \geq \varepsilon \|\mathbf{y}-\mathbf{x}\|^2.$$

d. Can you devise a computational scheme for solving the nonlinear complementary problem?

[11.14] Consider the following problem, where $\mathbf{A}$ is $m \times n$ and $\mathbf{H}$ is $n \times n$ and symmetric.

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

a. Write the KKT conditions.
b. Suppose that a point $\hat{\mathbf{x}}$ satisfies the KKT conditions. Is it necessarily true that $\hat{\mathbf{x}}$ is a global or local minimum?
c. Show that if $\mathbf{H}$ is positive semidefinite on the cone of feasible directions at $\hat{\mathbf{x}}$, then $\hat{\mathbf{x}}$ is a global optimal solution.

[11.15] This exercise describes a method credited to Dantzig [1963] for solving a quadratic programming problem of the form: Minimize $(1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, where $\mathbf{H}$ is symmetric and positive semidefinite. The KKT conditions for the above problem are

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} - \mathbf{v} &= \mathbf{0} \\ v_j x_j &= 0 \quad \text{for } j = 1, \ldots, n \\ \mathbf{x}, \mathbf{v} &\geq \mathbf{0}. \end{aligned}$$

The procedure always satisfies the first two conditions in addition to the nonnegativity of $\mathbf{x}$. The restriction $\mathbf{v} \geq \mathbf{0}$ is satisfied only at optimality. Furthermore, at each iteration, $v_j x_j = 0$ for all j except for at most one index.

Initialization Step Let $(\mathbf{x}_B^t, \mathbf{x}_N^t)$ be a basic feasible solution to $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, and let $\mathbf{v}^t = (\mathbf{v}_B^t, \mathbf{v}_N^t)$. Consider the basic solution to the system with the basic vectors $\mathbf{x}_B$, u, and $\mathbf{v}_N$. Note that the solution satisfies all the constraints except possibly $\mathbf{v} \geq \mathbf{0}$. Since u is unrestricted and since the algorithm relaxes $\mathbf{v} \geq \mathbf{0}$, as a variable enters the basis, only the x_j-variables are eligible to leave the basis.

Main Step

1. If $\mathbf{v} \geq \mathbf{0}$, stop. The current solution is optimal. Otherwise, let $v_j = \min \{v_i : v_i < 0\}$. Go to Step 2.
2. Introduce x_j into the basis. If v_j drops, go to Step 1. Otherwise, x_r drops for some r. Go to Step 3.
3. Introduce v_r into the basis. If v_j drops, go to Step 1. If another variable x_k drops, repeat Step 3 with v_r replaced by v_k.

a. Solve the following problem using the above method:

$$\begin{aligned} \text{Minimize} \quad & 3x_1^2 + 2x_2^2 - x_1x_2 \\ \text{subject to} \quad & -2x_1 + x_2 \leq 0 \\ & 2x_1 + 3x_2 \geq 6 \\ & 6x_1 + x_2 \leq 12 \\ & x_1, x_2 \geq 0. \end{aligned}$$

b. Prove that the above method converges to an optimal solution in a finite number of steps.

c. Consider the following problem credited to Finkbeiner and Kall [1973]:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + 3x_1 + 7x_3 + x_4 \\ \text{subject to} \quad & x_1 + 2x_2 + x_3 = 8 \\ & x_1 + 2x_2 + x_4 = 5 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

Starting with the basic variables $x_1 = 2$, $x_2 = 3$, $u_1 = 2$, $u_2 = 7$, $v_3 = 9$, and $v_4 = -6$, apply the above algorithm. Note that after one iteration, the variable v_1 should enter the basis but no appropriate variable could leave the basis, so that the method fails in the presence of linear terms in the objective function.

d. Consider the following modification in Step 3 of the above procedure suggested by Finkbeiner and Kall [1973]: If no variable drops from the basis when v_r is introduced, increase v_r if v_j does not decrease, or decrease v_r if v_j decreases without violating the nonnegativity of the **x** vector. Solve the problem in Part **c** by this method, and show that the procedure works in general.

[11.16] In this exercise we describe a procedure that is a modified version of a similar procedure credited to Wolfe [1959] for solving a quadratic programming problem of the form: Minimize $\mathbf{c}^t\mathbf{x}+(1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, where $\mathbf{A}$ is an $m \times n$ matrix of rank m and $\mathbf{H}$ is an $n \times n$ and symmetric matrix. The KKT conditions for this problem can be written as follows:

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b} \\ \mathbf{Hx}+\mathbf{A}^t\mathbf{u}-\mathbf{v} &= -\mathbf{c} \\ \mathbf{x},\mathbf{u} &\geq \mathbf{0} \\ \mathbf{v}^t\mathbf{x} &= 0. \end{aligned}$$

The method first finds a starting basic feasible solution to the system $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$. Using this solution, and denoting $\mathbf{A}$ by $[\mathbf{B}, \mathbf{N}]$ and $\mathbf{H}$ by $[\mathbf{H}_1, \mathbf{H}_2]$, where $\mathbf{B}$ is the basis, the above system can be rewritten as follows:

$$\begin{aligned} \mathbf{x}_B+\mathbf{B}^{-1}\mathbf{N}\mathbf{x}_N &= \mathbf{B}^{-1}\mathbf{b} \\ [\mathbf{H}_2-\mathbf{H}_1\mathbf{B}^{-1}\mathbf{N}]\mathbf{x}_N+\mathbf{A}^t\mathbf{u}-\mathbf{v} &= -\mathbf{H}_1\mathbf{B}^{-1}\mathbf{b}-\mathbf{c} \\ \mathbf{v}^t\mathbf{x} &= 0 \\ \mathbf{x}_B,\mathbf{x}_N,\mathbf{v} &\geq \mathbf{0}, \qquad \mathbf{u} \text{ unrestricted.} \end{aligned}$$

To start, we introduce n artificial variables in the last n constraints, with coefficient +1 if $(\mathbf{H}_1\mathbf{B}^{-1}\mathbf{b}+\mathbf{c})_i \leq 0$ and -1 if $(\mathbf{H}_1\mathbf{B}^{-1}\mathbf{b}+\mathbf{c})_i > 0$. We then have a basic feasible solution to the above system with the initial basis consisting of $\mathbf{x}_B$ and the artificial variables. The simplex method is then used to find a KKT point by minimizing the sum of the artificial variables. To maintain complementary slackness, the following restricted basis entry rule is adopted. If x_j is basic, then v_j cannot enter the basis, unless the minimum ratio test drives x_j out of the basis; conversely, if v_j is in the basis, then x_j cannot enter the basis unless the minimum ratio test drives v_j out of the basis.

a. What modifications are required if the constraint $\mathbf{Ax} = \mathbf{b}$ is replaced by $\mathbf{Ax} \leq \mathbf{b}$?

b. Use the above method to find a KKT point to the following quadratic program:

$$\begin{aligned} \text{Minimize} \quad & 3x_1^2 + 2x_1x_2 + 4x_2^2 - 3x_1 + 6x_2 \\ \text{subject to} \quad & 3x_1 + 2x_2 \le 6 \\ & x_1 + 3x_2 \le 6 \\ & x_1, x_2 \ge 0. \end{aligned}$$

c. Show that in the absence of degeneracy and under any of the following conditions, the above method produces a KKT point in a finite number of steps, assuming that the feasible region is non-empty.

(i) $\mathbf{H}$ is positive semidefinite and $\mathbf{c} = \mathbf{0}$.
(ii) $\mathbf{H}$ is positive definite.
(iii) $\mathbf{H}$ has nonnegative elements with strictly positive diagonal elements.

d. Show that if $\mathbf{H}$ is positive semidefinite and if, at termination, the sum of the artificial variables is not equal to zero, the quadratic program has an unbounded optimal solution.

e. Solve the following quadratic program by Wolfe's method:

$$\begin{aligned} \text{Minimize} \quad & -3x_1 - 5x_2 + 3x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{subject to} \quad & 2x_1 + 3x_2 \le 6 \\ & -x_1 + 2x_2 \le 2 \\ & x_1, x_2 \ge 0. \end{aligned}$$

[11.17] Consider the quadratic programming problem to minimize $\mathbf{c}^t\mathbf{x} + (1/)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} \le \mathbf{b}$, $\mathbf{x} \ge 0$, where $\mathbf{A}$ is an $m \times n$ matrix of rank m and $\mathbf{H}$ is $n \times n$ and symmetric. For simplicity, suppose that $\mathbf{b} \ge \mathbf{0}$. The KKT conditions can be written as follows:

$$\begin{aligned} \mathbf{A}\mathbf{x} + \mathbf{y} &= \mathbf{b} \\ -\mathbf{H}\mathbf{x} - \mathbf{A}^t\mathbf{u} + \mathbf{v} &= \mathbf{c} \\ \mathbf{v}^t\mathbf{x} = 0, \ \mathbf{u}^t\mathbf{y} &= 0 \\ \mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} &\ge \mathbf{0}. \end{aligned}$$

Now introduce the artificial variable z and consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & z \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} + \mathbf{y} && = \mathbf{b} \\ & -\mathbf{H}\mathbf{x} - \mathbf{A}^t\mathbf{u} + \mathbf{v} + \mathbf{q}z && = \mathbf{c} \\ & \mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} && \ge \mathbf{0}, \end{aligned}$$

where the ith component q_i of q is given by

$$q_i = \begin{cases} -1 & \text{if } c_i < 0 \\ 0 & \text{otherwise.} \end{cases}$$

We summarize below a modification of Wolfe's method described in Exercise 11.16 for solving the KKT system.

***Step* 1** Start with y and v as basic variables and note that some components of v may be negative. Let v_r be the most negative component of v. Pivot at the z column and the v_r row so that v_r is removed from the basis. We now have a basic solution with $z > 0$ and all variables nonnegative. Note that $x_j v_j = 0$, $\forall j = 1,\dots,n$ and that $u_i y_i = 0$, $\forall i = 1,\dots, m$.

***Step* 2** Minimize z by the simplex method using a restricted basis entry rule so that $v_j x_j = 0$ for $j = 1,\dots, n$ and $u_i y_i = 0$ for $i = 1,\dots, m$.

a. Solve the problem defined in Example 11.2.1 by the above procedure.
b. Suppose that **H** is positive semidefinite. Show that the above algorithm gives an optimal solution to the original problem or indicates that the problem is unbounded.
c. Show that if we delete the objective function row, the complementary pivoting algorithm discussed in Section 11.1 could be used to solve the KKT system. In this case, a variable enters the basis automatically if its complementary variable drops from the basis in the preceding iteration. Here x_j and v_j, and u_i and y_i, are complementary pairs of variables.

[11.18] Consider the quadratic program:

$$\text{QP: Minimize}\left\{\mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\right\},$$

where **A** is an $m \times n$ matrix of rank m and **H** is symmetric and positive definite on $\{\mathbf{x} : \mathbf{A}\mathbf{x} = 0\}$; that is, $\mathbf{x}^t\mathbf{H}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ satisfying $\mathbf{A}\mathbf{x} = \mathbf{0}$.

a. Show that the matrix $\begin{bmatrix} \mathbf{H} & \mathbf{A}^t \\ \mathbf{A} & \mathbf{0} \end{bmatrix}$ is nonsingular.
b. Hence, show that the linear equations of the KKT system for QP yield a unique solution.
c. Assuming that **H** is positive definite and hence nonsingular, derive an explicit closed-form expression for the optimal solution to QP.

[11.19] Consider the quadratic programming problem

$$\text{QP: Minimize}\left\{\mathbf{c}^t\mathbf{x}+\frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x}: \mathbf{A}_i^t\mathbf{x}=b_i \text{ for } i\in E,\ \mathbf{A}_i^t\mathbf{x}\le b_i \text{ for } i\in I\right\},$$

where $\mathbf{H}$ is symmetric and positive definite, and where the index sets E and I record the equality and inequality constraints in the problem, respectively. (Nonnegativities, if present, are included in the set indexed by I.) Consider the following *active set method* for solving QP. Given a feasible solution $\mathbf{x}_k$, define the working index set $W_k = E \cup I_k$, where $I_k \equiv \{i \in I : \mathbf{A}_i^t\mathbf{x}_k = b_i\}$ represents the binding inequality constraints, and consider the following direction-finding problem:

$$\text{QP}(\mathbf{x}_k)\text{: Minimize}\left\{(\mathbf{c}+\mathbf{H}\mathbf{x}_k)^t\mathbf{d}+\frac{1}{2}\mathbf{d}^t\mathbf{H}\mathbf{d}: \mathbf{A}_i^t\mathbf{d}=0 \text{ for all } i\in W_k\right\}.$$

Let d_k be the optimum obtained (see Exercise 11.18).

a. Show that $(\mathbf{x}_k+\mathbf{d}_k)$ solves the problem

$$\text{Minimize}\left\{\mathbf{c}^t\mathbf{x}+\frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x}: \mathbf{A}_i^t\mathbf{x}=b_i \text{ for all } i\in W_k\right\}. \tag{11.74}$$

b. If $\mathrm{d}_k = 0$, let v_i^*, $i \in W_k$, denote the optimum Lagrange multipliers for QP($\mathbf{x}_k$). If $v_i^* \ge 0$ for all $i \in I_k$, then show that $\mathbf{x}_k$ is optimal to QP. On the other hand, if the value $\min\{v_i^* : i\in I_k\} \equiv v_q^* < 0$, then let $I_{k+1} = I_k - \{q\}$, $W_{k+1} = E \cup I_{k+1}$, and $\mathbf{x}_{k+1} = \mathbf{x}_k$.

c. If $\mathbf{d}_k \ne \mathbf{0}$ and if $(\mathbf{x}_k+\mathbf{d}_k)$ is feasible to QP, put $\mathbf{x}_{k+1} = \mathbf{x}_k+\mathbf{d}_k$ and $W_{k+1} = W_k$. On the other hand, if $(\mathbf{x}_k+\mathbf{d}_k)$ is not feasible to QP, let $\alpha_k < 1$ be the maximum step length along $\mathbf{d}_k$ that maintains feasibility as given by

$$\alpha_k = \min_{i\notin I_k:\mathbf{A}_i^t\mathbf{d}_k>0}\left\{\frac{b_i-\mathbf{A}_i^t\mathbf{x}_k}{\mathbf{A}_i^t\mathbf{d}_k}\right\}=\frac{b_q-\mathbf{A}_q^t\mathbf{x}_k}{\mathbf{A}_q^t\mathbf{d}_k}.$$

Put $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$, $I_{k+1} = I_k \cup \{q\}$, and $W_{k+1} = E \cup I_{k+1}$. From Parts b and c, having determined $\mathbf{x}_{k+1}$ and W_{k+1}, increment k by 1 and reiterate. Let $\{\mathbf{x}_k\}$ be the sequence thus generated. Provide a finite convergence argument for this procedure by showing that while the solution to (11.74) is infeasible to QP, the method continues to add active constraints until the solution to (11.74) becomes feasible to QP and then either verifies optimality or provides a strict descent in the objective function value. Hence, using the fact that the number of possible working sets is finite, establish finite convergence of the procedure.

d. Illustrate by solving the problem of Example 11.2.1, starting at the origin.

[11.20] In this exercise we describe the method of Frank and Wolfe [1956] for solving a quadratic programming problem. This method generalizes a similar procedure by Barankin and Dorfman [1955]. Consider the problem to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} \le \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$, where $\mathbf{H}$ is symmetric and positive semidefinite.

a. Show that the KKT conditions can be stated as follows:

$$\begin{aligned} \mathbf{A}\mathbf{x}+\mathbf{x}_s &= \mathbf{b} \\ \mathbf{H}\mathbf{x}-\mathbf{u}+\mathbf{A}^t\mathbf{v} &= -\mathbf{c} \\ \mathbf{u}^t\mathbf{x}+\mathbf{v}^t\mathbf{x}_s &= 0 \\ \mathbf{x},\mathbf{x}_s,\mathbf{u},\mathbf{v} &\ge 0. \end{aligned}$$

The system can be rewritten as $\mathbf{E}\mathbf{y} = \mathbf{d}$, $\mathbf{y} \ge \mathbf{0}$, $\mathbf{y}^t\tilde{\mathbf{y}} = 0$, where

$$\mathbf{E} = \begin{bmatrix} \mathbf{A} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} & -\mathbf{I} & \mathbf{A}^t \end{bmatrix}, \quad \mathbf{d} = \begin{pmatrix} \mathbf{b} \\ -\mathbf{c} \end{pmatrix}$$

$$\mathbf{y}^t = (\mathbf{x}^t, \mathbf{x}_s^t, \mathbf{u}^t, \mathbf{v}^t)$$

$$\tilde{\mathbf{y}}^t = (\mathbf{u}^t, \mathbf{v}^t, \mathbf{x}^t, \mathbf{x}_s^t).$$

b. Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{y}^t\tilde{\mathbf{y}} \\ \text{subject to} \quad & \mathbf{E}\mathbf{y} = \mathbf{d} \\ & \mathbf{y} \ge \mathbf{0}. \end{aligned}$$

Show that a feasible point $\mathbf{y}$ satisfying $\mathbf{y}^t\tilde{\mathbf{y}} = 0$ yields a KKT point to the original problem.

c. Use the Frank-Wolfe method discussed in Exercise 10.15 to solve the problem stated in Part b, and show that the algorithm simplifies as follows. Suppose that at iteration k we have a basic feasible solution $\mathbf{y}_k$ for the above constraints and a feasible solution $\mathbf{w}_k$ to the same system, which is not necessarily basic. Starting with $\mathbf{y}_k$, solve the following linear program:

$$\begin{aligned} \text{Minimize} \quad & \tilde{\mathbf{w}}_k^t\mathbf{y} \\ \text{subject to} \quad & \mathbf{E}\mathbf{y} = \mathbf{d} \\ & \mathbf{y} \ge \mathbf{0}. \end{aligned}$$

A sequence of solutions is obtained ending with a point $\mathbf{y} = \mathbf{g}$, where either $\mathbf{g}^t\tilde{\mathbf{g}} = 0$ or $\mathbf{g}^t\tilde{\mathbf{w}}_k \le (1/2)\mathbf{w}_k^t\tilde{\mathbf{w}}_k$. In the former case, we stop with $\mathbf{g}$ as an optimal solution. In the latter case, set $\mathbf{y}_{k+1} = \mathbf{g}$ and let $\mathbf{w}_{k+1}$ be the convex combination of $\mathbf{w}_k$ and $\mathbf{y}_{k+1}$, which minimizes the objective function $\mathbf{y}^t\tilde{\mathbf{y}}$. Replace k by $k + 1$ and repeat the process. Show that this procedure converges to an optimal solution, and illustrate it by solving the following problem:

$$\begin{aligned} \text{Minimize} \quad & -3x_1 - 5x_2 + 2x_1^2 + x_2^2 \\ \text{subject to} \quad & 3x_1 + 2x_2 \le 6 \\ & 4x_1 + x_2 \le 4 \\ & x_1, x_2 \ge 0. \end{aligned}$$

d. Use the procedure of Frank and Wolfe described in Exercise 10.15 to solve the quadratic programming problem directly without first formulating the KKT conditions. Illustrate by solving the numerical problem in Part c, and compare the trajectories.

[11.21] In Section 11.2 we described a complementary pivoting procedure for solving a quadratic programming problem of the form to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$. We showed that the method produces an optimal solution if $\mathbf{H}$ is positive definite or if $\mathbf{H}$ is positive semidefinite and $\mathbf{c} \ge \mathbf{0}$. The following modification of the procedure to handle the case where $\mathbf{H}$ is positive semidefinite and $\mathbf{c} = \mathbf{0}$ is similar to that given by Wolfe [1959].

Step **1** Apply the complementary pivoting algorithm, where $\mathbf{c}$ is replaced by the zero vector. By Theorem 11.2.4 we obtain a complementary basic feasible solution to the following system:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} - \mathbf{v} &= \mathbf{0} \\ v_j x_j &= 0 \quad \text{for } j = 1, \ldots, n \\ \mathbf{x}, \mathbf{v} &\ge \mathbf{0}, \ \mathbf{u} \text{ unrestricted.} \end{aligned}$$

Step **2** Starting with the solution obtained in Step 1, solve the following problem using the simplex method with the restricted basis entry rule so that v_j and x_j are never in the basis simultaneously:

$$\begin{aligned} \text{Maximize} \quad & z \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} - \mathbf{v} + z\mathbf{c} = \mathbf{0} \\ & \mathbf{x} \ge \mathbf{0}, \mathbf{v} \ge \mathbf{0}, z \ge 0, \qquad \mathbf{u} \text{ unrestricted.} \end{aligned}$$

At optimality, either $\bar{z} = 0$ or $\bar{z} = \infty$ along an extreme direction. In the former case, the optimal solution to the quadratic program is unbounded. In the latter case, an optimal solution to the quadratic program is determined by letting $z = 1$ along the ray giving rise to the unbounded solution.

a. Show that if the optimal objective value to the problem in Step 2 is finite, then it must be zero. Show in this case that the optimal objective value of the original problem is unbounded.
b. Show that if the optimal objective value $\bar{z} = \infty$, then the solution along the optimal ray with $z = 1$ still maintains complementary slackness and, hence, gives an optimal solution to the original problem.
c. Solve the problem of Example 11.2.1 by the above procedure.

[11.22] Let $f(\mathbf{x}) = \mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$, where $\mathbf{H}$ is a symmetric positive semidefinite matrix. Show that $f: R^n \to R$ is *unbounded* from below if and only if $\mathbf{c} + \mathbf{H}\mathbf{x} = \mathbf{0}$ has no solution.

[11.23] Consider the RLT linear programming relaxation given in Equation (11.17).

a. Show that if some inequality $\mathbf{G}_r\mathbf{x} \le g_r$ is implied by the remaining inequalities $\mathbf{G}_i\mathbf{x} \le g_i$, $i = 1,\ldots, \bar{m}$, $i \ne r$, then any RLT constraint of the type (11.17b) that is generated via a pair that includes the factor $(g_r - \mathbf{G}_r\mathbf{x}) \ge 0$ is implied by the RLT constraints (11.17b) generated via pairwise products of the remaining constraints.
b. For Example 11.2.8, show that the restriction $x_2 \le 15$ is implied by the remaining defining inequalities. Hence, verify Part a with respect to this inequality, and specifically identify the 15 RLT inequalities (11.17b) that you would generate for LP(Ω) by omitting the factor $(15 - x_2) \ge 0$ from the pairwise products. Verify that your resulting LP relaxation obtains the same objective value as that derived for LP(Ω) in Example 11.2.8.

[11.24] Provide a complete formulation for each of the LP relaxations LP(Ω^q), $q = 1, 2, 3$, as identified in Example 11.2.8, and solve these relaxations. Hence, verify the optimality of the solutions (0, 6) and (24, 6) for the underlying quadratic program.

[11.25] Consider the quadratic program of Example 11.2.8. Defining additional *cubic RLT variables* $w_{111} = x_1^3$, $w_{112} = x_1^2x_2$, and $w_{122} = x_1x_2^2$, consider the following linear programming relaxation, having selected second- and third-order RLT constraints, where $s_1 \equiv 48 + 6x_1 - 8x_2 \ge 0$ and $s_2 \equiv 120 - 3x_1 - 8x_2 \ge 0$ represent the constraint factors.

$$\text{LP: Minimize } \{-w_{11} - w_{22} + 24x_1 - 144 : [(24 - x_1)s_1]_L \geq 0, [x_1, s_2]_L \geq 0, \\ [(24 - x_1)x_1]_L \geq 0, [(24 - x_1)x_1 s_1]_L \geq 0, \\ [(24 - x_1)x_2 s_1]_L \geq 0, [(24 - x_1)x_1 s_2]_L \geq 0, \\ \text{and} [x_1 x_2 s_2]_L \geq 0.$$

a. Verify that this LP yields an optimal solution given by $(\hat{x}_1, \hat{x}_2) = (0, 6)$, $\hat{w}_{22} = 36$, and $\hat{w}_{11} = \hat{w}_{12} = \hat{w}_{111} = \hat{w}_{112} = \hat{w}_{122} = 0$, having the objective function value of -180. Hence, argue why this single LP solution leads directly to the conclusion that $(x_1, x_2) = (0, 6)$ solves the underlying quadratic program.

b. Use Exercise 3.32 to construct the *convex envelope* of the concave objective function of the quadratic program given in Example 11.2.8 over the polytope that defines its feasible region. In particular, letting z represent the objective function, show that $z \geq -6x_2 - 144$ and $z \geq (10/3)x_2 - 200$ defines two principal constraints representing the epigraph of this convex envelope.

c. Show that an appropriate surrogate of the constraints of LP in Part a along with the objective representation $z = -w_{11} - w_{22} + 24x_1 - 144$ reproduces the two epigraph-defining constraints given in Part b.

[11.26] Consider the RLT linear programming relaxation given by Equation (11.17) for the underlying quadratic programming problem. Let **W** denote the $n \times n$ symmetric matrix having $w_{(ij)}$ as its (i, j)th element.

a. Verify that **W** represents the *diadic* or *outer product* $[\mathbf{x}\mathbf{x}^t]_L$, where $[\cdot]_L$ denotes the linearization under the substitution (11.16). Hence, show that it is valid to include the restrictions

$$\mathbf{X} \equiv \begin{bmatrix} \mathbf{W} & \mathbf{x} \\ \mathbf{x}^t & 1 \end{bmatrix} \succeq 0$$

within the relaxation LP(Ω) given by (11.17), where $\succeq 0$ denotes symmetric and positive semidefinite. (This yields a *semidefinite programming* relaxation for NQP.)

b. Explore the literature for possible solution approaches to the resulting semidefinite program defined in Part a (see the Notes and References section).

c. Show that $\mathbf{X} \succeq 0$ can be replaced by the *RLT constraints* (in addition to **X** being symmetric) that $\alpha^t \mathbf{X} \alpha \geq 0$, $\forall\, \alpha \in R^{n+1}$, such that $\|\alpha\| = 1$.

d. Suppose now that we solve the LP(Ω) relaxation (11.17) and obtain a solution $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$, and we use this solution to compose the matrix $\bar{\mathbf{X}}$. Show using the superdiagonization algorithm of Chapter 3 how you can either verify in polynomial time that $\overline{\mathbf{X}} \succeq 0$, or else generate an $\overline{\alpha} \in R^{n+1}$ of the type in Part c for which $\overline{\alpha}^t \overline{\mathbf{X}} \overline{\alpha} < 0$. Hence, show that it is valid to impose the *semidefinite cut* $\overline{\alpha}^t \overline{\mathbf{X}} \overline{\alpha} \geq 0$ within LP(Ω).

e. Devise a revised enhanced algorithm for solving Problem NQP of Section 11.2 that incorporates the use of suitable semidefinite cuts from Part d. (See Sherali and Fraticelli [2003] for one such type of algorithm.)

[11.27] Consider the *polynomial programming problem*

$$\text{PP: Minimize}\{\phi_0(\mathbf{x}) : \phi_r(\mathbf{x}) \geq \beta_r,\ r = 1,\dots,m,\ \mathbf{x} \in \Omega\},$$

where $\Omega = \{\mathbf{x} : 0 \leq \ell_j \leq x_j \leq u_j < \infty \text{ for } j = 1,\dots, n\}$ and where $\phi_r(\mathbf{x}) \equiv \sum_{t \in T_r} \alpha_{rt} \prod_{j \in J_{rt}} x_j$ for $r = 0, 1,\dots, m$. Hence, for each $r = 0, 1,\dots, m$, T_r is an index set for the terms defining the *polynomial function* ϕ_r, α_{rt} are real coefficients for the *multinomial terms* $(\prod_{j \in J_{rt}} x_j)$, $\forall\, t \in T_r$, and where J_{rt} is a *multiset* that contains possibly repeated elements from $N \equiv \{1,\dots, n\}$. (For example, if $J_{rt} = \{1, 2, 2, 3\}$, then the corresponding multinomial term is $x_1 x_2^2 x_3$.) In particular, let δ denote the highest degree of any polynomial function defining the problem, and accordingly, let $\overline{N} \equiv \{N,\dots, N\}$ represent δ copies of N. Hence, each $J_{rt} \subseteq \overline{N}$ with $1 \leq |J_{rt}| \leq \delta$, $\forall\, t \in T_r$, $r = 0, 1,\dots, R$.

Consider the following *reformulation-linearization/convexification technique* (RLT)-based process. Include within PP the following linearized *bound-factor product* constraints:

$$\left[\prod_{j \in J_1} (x_j - \ell_j) \prod_{j \in J_2} (u_j - x_j)\right]_L \geq 0,\ \forall\, J_1 \cup J_2 \subseteq \overline{N} \text{ and } |J_1 \cup J_2| = \delta, \tag{11.75}$$

where $[\cdot]_L$ denotes the linearization of $[\cdot]$ under the substitution

$$X_J = \prod_{j \in J} x_j,\ \forall\, J \subseteq \overline{N},\ 2 \leq |J| \leq \delta, \tag{11.76}$$

and where the indices in J are assumed to be sequenced in nondecreasing order. Let $\mathbf{X}$ denote the vector of variables $(X_J,\ J \subseteq \overline{N},\ 2 \leq |J| \leq \delta)$.

a. Verify that there are $\sum_{k=0}^{\delta} \binom{n+k-1}{k} \binom{n+(\delta-k)-1}{\delta-k}$ inequalities of the type (11.75).

b. Let LP(Ω): $\text{Minimize}\{[\phi_0(\mathbf{x})]_L : [\phi_r(\mathbf{x})]_L \geq \beta_r,\ r = 1,\dots, R,\ x \in \Omega,$ plus the inequalities (11.75)}. (11.77)

Show that $v[\mathrm{LP}(\Omega)] \le v[\mathrm{PP}]$. Moreover, show that if $(\bar{\mathbf{x}}, \bar{\mathbf{X}})$ solves $\mathrm{LP}(\Omega)$ and satisfies (11.76), then $\bar{\mathbf{x}}$ solves PP.

c. Let $(\bar{\mathbf{x}}, \bar{\mathbf{X}})$ be any feasible solution to $\mathrm{LP}(\Omega)$. Suppose that $\bar{x}_p = \ell_p$ or $\bar{x}_p = u_p$ for some $p \in \{1,..., n\}$. Then show that

$$\bar{X}_{(J \cup \{p\})} = \bar{x}_p \bar{X}_J, \ \forall\, J \subseteq \bar{N},\ 1 \le |J| \le \delta - 1$$

when $X_{\{j\}} \equiv x_j, \forall j$. Interpret this result in light of Lemma 11.2.6.

d. Design a branch-and-bound algorithm to solve Problem PP, similar to the RLT algorithm described in Section 11.2 for Problem NQP, based on the results of Parts b and c. Also state and prove a convergence result similar to that of Theorem 11.2.7. (This development is due to Sherali and Tuncbilek [1992].)

[11.28] Consider the quadratic programming problem to minimize $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} \le \mathbf{b}$, where $\mathbf{H}$ is an $n \times n$ symmetric positive definite matrix and $\mathbf{A}$ is an $m \times n$ matrix. For any subset $S \subseteq \{1,..., m\}$ of the constraint indices, let $\mathbf{x}_s$ be a minimum of $\mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{H}\mathbf{x}$ subject to the constraints in S as binding, and let $V(\mathbf{x}_s)$ be the set of constraints violated by $\mathbf{x}_s$.

a. Show that if $V(\mathbf{x}_s) \ne \varnothing$, then S could be a subset of the set of binding constraints $\hat{S}$ at optimality only if there exists some $h \in \hat{S} \cap V(\mathbf{x}_s)$.

b. Show that if $V(\mathbf{x}_s) = \varnothing$, then $\mathbf{x}_s$ is an optimal solution to the original problem if and only if $h \in V(\mathbf{x}_{s-h})$ for each $h \in S$.

c. From Parts a and b, show that the following *active set strategy* credited to Theil and van de Panne [1961] solves the quadratic problem. First, solve the unconstrained problem so that $S = \varnothing$. If $V(\mathbf{x}_\varnothing) = \varnothing$, then $\mathbf{x}_\varnothing$ is an optimal solution. Otherwise, form sets of the type S_1, where $S_1 = \{h\}$ and $h \in V(\mathbf{x}_\varnothing)$. Find $\mathbf{x}_{s_1}$ for each such S_1. If $V(\mathbf{x}_{s_1}) = \varnothing$ for some S_1, check by Part b whether $\mathbf{x}_{s_1}$ is optimal. If no candidate problems of the form S_1 could produce an optimal solution, form sets of the type S_2 with two binding constraints, where $S_2 = S_1 \cup \{h\}$ and where S_1 is a set having one binding constraint such that $V(\mathbf{x}_{s_1}) \ne \varnothing$ and $h \in V(\mathbf{x}_{s_1})$. The process is repeated by finding $\mathbf{x}_{s_2}$ that solves the original problem or by forming sets of the type S_3 containing three binding constraints.

d. Illustrate the method of Theil and van de Panne by solving the problem stated in Example 11.2.1.

e. Can you generalize the method to the following convex programming problem where f is strictly convex and g_i is convex for $i = 1,\ldots, m$?

$$\begin{aligned} &\text{Minimize} \quad f(\mathbf{x}) \\ &\text{subject to} \quad g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1,\ldots,m. \end{aligned}$$

[11.29] Does Theorem 11.3.1 hold true if $f_1,\ldots, f_n$ are convex rather than strictly convex? If not, modify the statement of the theorem so that it handles the convex case.

[11.30] Solve the following problem using the method discussed in Section 11.3:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2x_1+1} + 3x_2^3 \\ &\text{subject to} \quad 2x_1^2 - x_2^3 \le 4 \\ &\qquad\qquad x_1, x_2 \ge 6. \end{aligned}$$

[11.31] Consider the following problem:

$$\begin{aligned} &\text{Minimize} \quad -2x_1 + 3x_2 + 3x_1^2 - 2x_1x_2 + 2x_2^2 \\ &\text{subject to} \quad 2x_1 + x_2 \le 6 \\ &\qquad\qquad x_1^2 + x_2^2 = 9 \\ &\qquad\qquad x_1, x_2 \ge 0. \end{aligned}$$

Make a suitable change of variables such that the problem becomes separable. Choose suitable grids for partitioning, set up the initial simplex tableau, and then solve the approximating problem. If you were to solve the problem over again, how can you make use of your answer to obtain a better partitioning?

[11.32] Consider the following problem:

$$\begin{aligned} &\text{Minimize} \quad 3e^{2x_1} + 2x_1^2 + 3x_1 + 2x_2^2 - 5x_2 + 3x_3 \\ &\text{subject to} \quad x_1^2 + e^{x_2} + 6x_3 \le 15 \\ &\qquad\qquad x_1^4 - x_2 + 5x_3 \le 25 \\ &\qquad\qquad 0 \le x_1 \le 4 \\ &\qquad\qquad 0 \le x_2 \le 2 \\ &\qquad\qquad 0 \le x_3 \le \infty. \end{aligned}$$

a. Using the grid points 0, 2, and 4 for x_1, and 0, 1, and 2 for x_2, solve the above problem by the separable programming algorithm.
b. Starting from the optimal solution obtained in Part a, use the grid point generation scheme to generate three more grid points to obtain a better solution.

c. Using the optimal point obtained in Part b, give lower and upper bounds on the optimal objective value to the original problem.

[11.33] Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & 2e^{x_1} + e^{3x_2} + x_1 + 2x_2^2 + 3x_1^2 \\ \text{subject to} \quad & 3x_1 + 2x_2 \le 6 \\ & 2x_1 - x_2 \le 0 \\ & x_1, x_2 \ge 0. \end{aligned}$$

a. Show that the objective function is strictly convex and that the constraints are convex. Hence, the restricted basis entry can be dropped if the problem is to be solved by the separable programming method discussed in Section 11.3.
b. Use suitable grid points and solve the problem.

[11.34] Does the simplex method with the restricted basis entry rule provide an optimal solution to the approximating Problem LAP in the nonconvex case? Prove or give a counterexample.

[11.35] Consider the following alternative method for approximating a function θ in the interval $[a, b]$. The interval $[a, b]$ is divided into smaller subintervals via the grid points $a = \mu_1, \ldots, \mu_k = b$. Let $\Delta_i = \mu_{i+1} - \mu_i$, and let $\Delta\theta_i = \theta(\mu_{i+1}) - \theta(\mu_i)$ for $i = 1, \ldots, k-1$. Now, consider x in the interval $[\mu_\nu, \mu_{\nu+1}]$. Then x can be represented as $x = \mu_1 + \sum_{i=1}^{\nu} \delta_i \Delta_i$, and $\theta(x)$ can be approximated by $\hat{\theta}(x) = \theta_1 + \sum_{i=1}^{\nu} \delta_i \Delta\theta_i$, where $\delta_\nu \in [0, 1]$, $\delta_i = 1$ for $i = 1, \ldots, \nu - 1$, and $\theta_1 = \theta(x_1)$.

a. Interpret this approximation of θ geometrically.
b. Show how this approximation can be used to solve the following separable problem by the simplex method, with a suitable restricted basis entry rule:

$$\begin{aligned} \text{Minimize} \quad & \sum_{j=1}^{n} f_j(x_j) \\ \text{subject to} \quad & \sum_{j=1}^{n} g_{ij}(x_j) \le 0 && \text{for } i = 1, \ldots, m \\ & a_j \le x_j \le b_j && \text{for } j = 1, \ldots, n. \end{aligned}$$

[*Hint*: Let $x_{\nu j}$ for $\nu = 1, \ldots, k_j + 1$ be the grid points used for x_j, and consider the following problem:

$$\text{Minimize} \quad \sum_{j=1}^{n} \sum_{v=1}^{k_j} (\Delta f_{vj})\delta_{vj} + \sum_{j=1}^{n} f_j(a_j)$$

$$\text{subject to} \quad \sum_{j=1}^{n} \sum_{v=1}^{k_j} (\Delta g_{ijv})\delta_{vj} + \sum_{j=1}^{n} g_{ij}(a_j) \le 0 \qquad \text{for } i = 1,\ldots,m$$

$$0 \le \delta_{vj} \le 1 \qquad v = 1,\ldots,k_j;\ j = 1,\ldots, n$$

$$\delta_{vj} > 0 \Rightarrow \delta_{\ell j} = 1 \qquad \text{for } \ell < v;\ j = 1,\ldots, n$$

where

$$\Delta f_{vj} = f(x_{v+1,j}) - f(x_{vj})$$

$$\Delta g_{ijv} = g_j(x_{v+1,j}) - g_{ij}(x_{vj}).]$$

c. Use the procedure developed in Part b to solve the following problem:

$$\begin{aligned} \text{Maximize} \quad & 3x_1 + 4x_2 - 3x_1^2 - 2x_2^2 \\ \text{subject to} \quad & 2x_1 + 3x_2 \le 12 \\ & -2x_1 + 3x_2 \le 6 \\ & x_1, x_2 \ge 0. \end{aligned}$$

[11.36] In Section 11.3 we approximated a separable programming problem using the λ-form. An alternative, called the *δ-form approximation* was considered in Exercise 11.35. Consider a variable x in the interval $[a, b]$ and grid points $\mu_1 = a, \mu_2,\ldots,\mu_k = b$. Then, using the λ and δ forms, x can be represented, respectively by:

1. $x = \sum_{j=1}^{k} \lambda_j \mu_j$, $\sum_{j=1}^{k} \lambda_j = 1$, $\lambda_j \ge 0$, for $j = 1,\ldots, k$, where $\lambda_p \lambda_q = 0$ if μ_p and μ_q are not adjacent.

2. $x = \mu_1 + \sum_{j=1}^{k-1} \Delta_j \delta_j$, $0 \le \delta_j \le 1$ for $j = 1,\ldots, k-1$, and $\delta_i > 0 \Rightarrow \delta_j = 1$ for $j < i$.

Show that the two forms are related by the relationship

$$\lambda_j = \begin{cases} \delta_{j-1} - \delta_j & \text{if } j = 1,\ldots,k-1 \\ \delta_{j-1} & \text{if } j = k, \end{cases}$$

where $\delta_0 = 1$. In particular, show that this relationship could be written in vector form as $\boldsymbol{\lambda} = \mathbf{T}\boldsymbol{\delta}$, where T is an upper triangular matrix.

[11.37] Consider the function f defined by

$$f(\mathbf{x}) = \frac{x_1 + 2x_2 - 6}{3x_1 - x_2 + 2}.$$

a. Sketch the following sets in the (x_1, x_2)-plane and determine whether they are convex:

$$S = \{(x_1, x_2) : f(\mathbf{x}) \le 2\}$$
$$S_1 = S \cap \{(x_1, x_2) : 3x_1 - x_2 + 2 > 0\}$$
$$S_2 = S \cap \{(x_1, x_2) : 3x_1 - x_2 + 2 < 0\}.$$

b. Is your conclusion in Part a inconsistent with the fact that f is quasiconvex on the region $\{(x_1, x_2) : 3x_1 - x_2 + 2 \ne 0\}$? Discuss.

[11.38] Consider the following problem:

$$\begin{aligned} \text{Maximize} \quad & \frac{7x_1 + 5x_2 - 3}{-4x_1 + 2x_2 - 40} \\ \text{subject to} \quad & x_1 + x_2 \le 10 \\ & 3x_1 - 5x_2 \le 6 \\ & x_1, x_2 \ge 0. \end{aligned}$$

a. Solve the problem by the method of Gilmore and Gomory.
b. Solve the problem by the method of Charnes and Cooper.

[11.39] Solve the following problem by the two linear fractional programming algorithms discussed in Section 11.4:

$$\begin{aligned} \text{Minimize} \quad & \frac{-3x_1 + 2x_2 + 4x_3 + 3}{2x_1 + x_2 + 3x_3 + 2} \\ \text{subject to} \quad & 3x_1 + 2x_2 + 4x_3 \le 12 \\ & 2x_1 + x_2 \ge 2 \\ & x_1 + 3x_3 \le 8 \\ & x_1, x_2, x_3 \ge 0. \end{aligned}$$

[11.40] Let

$$f(\mathbf{x}) = \frac{\mathbf{p}^t\mathbf{x} + \alpha}{\mathbf{q}^t\mathbf{x} + \beta}$$

and let $S = \{\mathbf{x} : \mathbf{q}^t\mathbf{x} + \beta > 0\}$. Show directly that f is quasiconvex, quasiconcave, strictly quasiconvex, and strictly quasiconcave on S.

[11.41] Suppose that the region $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \ge \mathbf{0}\}$ is unbounded. Further, suppose that an improving feasible direction $\mathbf{d}$ is found while minimizing a linear fractional function over the above region. In particular, suppose that $\mathbf{d}_N$

consists of a vector of zeros except for a 1 at position j and $\mathbf{d}_B = -\mathbf{B}^{-1}\mathbf{a}_j \geq \mathbf{0}$. Is it necessarily true that the optimal objective is unbounded by moving from the current extreme point in the direction $\mathbf{d}$? If not, what possible cases can be encountered?

[11.42] Let f: $R^n \to R$ be quasiconcave, and let $\theta(\lambda) = f(\mathbf{x}+\lambda\mathbf{d})$, where $\mathbf{x}$ is a given vector and $\mathbf{d}$ is a given direction.

a. Show that θ is a quasiconcave function in λ.
b. Consider the problem to minimize $\theta(\lambda)$ subject to $\lambda \in [a, b]$. Show that if $\nabla f(\mathbf{x})^t\mathbf{d} < 0$, then $\lambda = b$ is an optimal solution to the above problem.
c. Letting $f(\mathbf{x}) = (\mathbf{p}^t\mathbf{x}+\alpha)/(\mathbf{q}^t\mathbf{x}+\beta)$, use the result in Part b to show that no line search is needed for solving linear fractional programs by the convex-simplex method.

[11.43] In solving a linear fractional programming problem, suppose that we add the following two rows to the initial tableau:

$$z_1 - \mathbf{p}^t\mathbf{x} = \alpha$$

$$z_2 - \mathbf{q}^t\mathbf{x} = \beta.$$

As the problem is solved by the convex-simplex method, the coefficients of the basic vector $\mathbf{x}_B$ in these rows are equal to zero, so that the updated rows are given by

$$z_1 - (\mathbf{p}_N^t - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{N})\mathbf{x}_N = \alpha + \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{b}$$

$$z_2 - (\mathbf{q}_N^t - \mathbf{q}_B^t\mathbf{B}^{-1}\mathbf{N})\mathbf{x}_N = \beta + \mathbf{q}_B^t\mathbf{B}^{-1}\mathbf{b}.$$

Show that the reduced gradient vector $\mathbf{r}_N$ is given by

$$\mathbf{r}_N = \frac{(\mathbf{p}_N^t - \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{N})\bar{z}_2 - (\mathbf{q}_N^t - \mathbf{q}_B^t\mathbf{B}^{-1}\mathbf{N})\bar{z}_1}{\bar{z}_2^2},$$

where $\bar{z}_1 = \alpha + \mathbf{p}_B^t\mathbf{B}^{-1}\mathbf{b}$ and $\bar{z}_2 = \beta + \mathbf{q}_B^t\mathbf{B}^{-1}\mathbf{b}$. Note that each term in the expression for $\mathbf{r}_N$ is immediately available from the updated tableau. Solve the problem in Example 11.4.3 using the above procedure for computing $\mathbf{r}_N$.

[11.44] Verify that the separable objective function (11.62a) of the dual geometric program (DGP) is concave.

[11.45] Consider the geometric programming problem of Example 11.5.4, and let C denote the ratio of the cost per unit length (cm) of the wire to the cost per unit surface area (cm^2) of the cylinder. Analyze this problem to study the sensitivity of the optimal dimensions of the cylinder to this cost factor C.

[11.46] Consider the geometric programming problem to

$$\begin{aligned} \text{Minimize} \quad & 35x_1^2x_2 + 15x_2x_3 \\ \text{subject to} \quad & \frac{2}{5}x_1^{-1}x_2^{-1/3} + \frac{3}{5}x_2^{-2}x_3^{-4/3} \le 1 \\ & \mathbf{x} > \mathbf{0}. \end{aligned}$$

State the degree of difficulty of this problem and solve it.

[11.47] Consider the problem to minimize $f_1(\mathbf{x}) + [f_2(\mathbf{x})]^a f_3(\mathbf{x})$, where f_i, $i = 1, 2, 3$, are posynomials and where $a > 0$. Show that this is equivalent to the standard posynomial geometric program GP to minimize $f_1(\mathbf{x}) + x_0^a f_3(\mathbf{x})$ subject to $x_0^{-1} f_2(\mathbf{x}) \le 1$, where x_0 is an additional variable. Illustrate by solving the problem to minimize $2x_1^{-1/3}x_2^{1/6} + [(3/5)x_1^{1/2}x_2^{3/4} + (2/5)x_1^{2/3}x_2]^{1/2} x_1^{3/4}x_2^{-1/3}$.

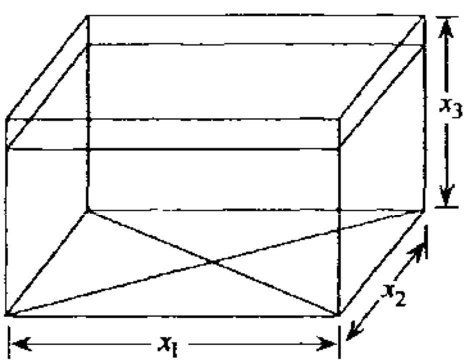
[11.48] Consider the geometric programming problem to

$$\begin{aligned} \text{Minimize} \quad & 25x_1^{-2}x_2^{-1/2}x_3^{-1} + 20x_1^2x_3 + 30x_1x_2^2x_3 \\ \text{subject to} \quad & \frac{5}{3}x_1^{-1}x_2^{-2} + \frac{4}{3}x_2^{1/2}x_3^{-2} \le 1 \\ & \mathbf{x} > \mathbf{0}. \end{aligned}$$

State the degree of difficulty of this problem and solve it.

[11.49] Re-solve the geometric programming problem of Example 11.5.3, assuming that the cylinder is open at one end.

[11.50] Suppose that a metal wire frame has to be constructed for a rectangular box having a skeleton and dimensions (in centimeters) as shown below.



Formulate the problem of minimizing the total length of wire used subject to the volume being at least 15 cm^3 as a standard posynomial geometric programming problem. What is the degree of difficulty of your formulation? Solve this problem.

[11.51] Consider the problem to minimize $f_1(\mathbf{x}) - f_2(\mathbf{x})$, where f_1 and f_2 are posynomials, f_2 has only one term, and the optimal value is known to be negative. Show that this can equivalently be solved as the standard posynomial geometric program to minimize x_0^{-1} subject to $[x_0/f_2(\mathbf{x})] + [f_1(\mathbf{x})/f_2(\mathbf{x})] \le 1$.

[11.52] Consider the problem to minimize

$$f_1(\mathbf{x}) + \frac{f_2(\mathbf{x})}{[f_3(\mathbf{x}) - f_4(\mathbf{x})]^a},$$

where f_i, $i = 1,\ldots,4$, are posynomials but f_3 has only one term, and where $a > 0$. Show that this is equivalent to the standard posynomial geometric program to minimize $f_1(\mathbf{x}) + f_2(\mathbf{x})x_0^{-a}$ subject to $x_0/f_3(\mathbf{x}) + f_4(\mathbf{x})/f_3(\mathbf{x}) \le 1$.

[11.53] Referring to Exercise 11.47, solve the problem to minimize $x_1^{-1/2}x_2^{1/8} + [(4/5)x_1^{1/2}x_2^{2/3} + (2/5)x_1^{1/3}x_2]^{1/2}x_1^{1/4}x_2^{-1/2}$.

[11.54] Consider the geometric programming problem to

$$\begin{aligned} \text{Minimize} \quad & 40x_1x_2 + 20x_2x_3 \\ \text{subject to} \quad & \frac{1}{5}x_1^{-1}x_2^{-1/2} + \frac{3}{5}x_2^{-1}x_3^{-2/3} \le 1 \\ & \mathbf{x} > \mathbf{0}. \end{aligned}$$

State the degree of difficulty of this problem and solve it.

[11.55] Consider the geometric programming problem to

$$\begin{aligned} \text{Minimize} \quad & 40x_1^{-1}x_2^{-1/2}x_3^{-1} + 20x_1x_3 + 40x_1x_2x_3 \\ \text{subject to} \quad & \frac{1}{3}x_1^{-2}x_2^{-2} + \frac{4}{3}x_2^{1/2}x_3^{-1} \le 1 \\ & \mathbf{x} > \mathbf{0}. \end{aligned}$$

State the degree of difficulty of this problem and solve it.

Notes and References

In Section 11.1 we introduced the linear complementary problem. The KKT optimality conditions for linear and quadratic programs can be expressed as linear complementary problems. The problem also arises in several contexts, such as bimatrix games and engineering optimization. The interested reader may refer to Cottle and Dantzig [1968], Dennis [1959], Du Val [1940], Kilmister and Reeve [1966], Lemke [1965, 1968], Lemke and Howson [1964], and Murty [1976, 1988]. LCPs also arise in finite difference schemes for fractional value problems (Cryer [1971]) and in electronic circuit stimulation problems (Van Bokhoven [1980]). The general LCP has been shown to be NP-complete in the strong sense, so we may not expect even a pseudo-polynomial time algorithm unless P = NP (see Garey and Johnson [1979]). In 1968, Lemke proposed a complementary pivoting algorithm, which is discussed in Section 11.1, for solving the linear complementary problem. Lemke proved that the method converges in a finite number of steps to a complementary basic feasible solution to the system if the matrix **M** is copositive plus. Eaves [1971a,b] extended the result to a more general class of matrices. In 1974, van de Panne developed a variant of Lemke's method for solving the linear complementary problem. Mangasarian [1976] and Solow and Sengupta [1985] give other optimization-based approaches, Kostreva [1978] provides an algebraic approach, and Cottle and Pang [1978] give a topological scheme. Cottle [1968] and Cottle and Dantzig [1968] have designed a *principal pivoting algorithm*, which was introduced in Exercise 11.11. In particular, if **M** is positive semidefinite, or if **M** is a *P*-matrix (having all principal minors positive), Murty [1972] has shown that the principal pivoting methods of Cottle and Dantzig [1968] or Lemke's algorithm [1968] can solve this problem. Also, see Murty [1988], and Rohn [1990] for further discussion of principal pivoting methods. Todd [1974] presents a general pivoting system that provides a natural setting for the study of complementary pivoting algorithms.

When **M** is positive semidefinite, Chung and Murty [1981] and Kozlov et al. [1979] have shown that Khachiyan's [1979a,b] algorithm can be modified to solve LCP polynomially. Cottle and Veinott [1969], Mangasarian [1979], Kojima et al. [1991], and Ye [1988] also discuss polynomially solvable cases, where **M** is a *Z*-matrix, or is positive semidefinite, or is skew-symmetric, or belongs to a restricted class of *P*-matrices. Ye and Pardalos [1989] employ an interior point potential reduction algorithm to develop another class of LCPs that are polynomially solvable, suggesting that positive semidefiniteness of **M** may not be the fundamental demarcation between P and NP in the context of LCPs. Several attempts have also been made to solve LCPs as a single linear program (Mangasarian [1976, 1978, 1979]) or as a sequence of linear programs (see Roy and Solow [1985] and Shiau [1983]).

Al-Khayyal [1987] develops a branch-and-bound algorithm for solving general LCPs. Pardalos and Rosen [1988] (see Exercise 11.6) show how LCPs can be formulated as mixed-integer linear programming problems, and develop an efficient heuristic. Al-Khayyal [1990] also shows how LCPs can be solved as bilinear programming problems (see Al-Khayyal and Falk [1983], Konno and Yajima [1989], Sherali and Alameddine [1992], Sherali and Shetty [1980], and Vaish and Shetty [1976, 1977] for competitive approaches). Concave minimization approaches (Bard and Falk [1982], Sherali et al. [1996]) and other linear mixed-integer zero-one and cutting plane approaches (see Sherali et al. [1998] and Vandenbussche and Nemhauser [2003]) have also been suggested. (See Exercises 11.4 and 11.9.) (See Parker and Rardin [1988] and Nemhauser and Wolsey [1999] for a discussion on solving mixed-integer problems.) Kostreva and Wiecek [1989] draw interesting interrelationships between LCPs and multiple objective programming problems.

The linear complementary problem has been extended to the nonlinear case and was introduced briefly in Exercise 11.13. The KKT conditions for a general nonlinear programming problem can be expressed as a nonlinear complementary problem. There has been a considerable amount of research on the existence of solutions to such a problem, but little has been done in the area of development of computational schemes for finding such solutions. See Cottle [1966], Eaves [1971b], Habetler and Price [1971, 1973], Karamardian [1969, 1971, 1972], and Murty [1988].

There are several approaches for solving quadratic programming problems. The methods of feasible directions discussed in Chapter 10 could be used to solve the problem. One such implementation is Beale's method [1955, 1959], which is essentially a specialization of the convex-simplex method. Another popular procedure that has been used is a combinatorial approach to iteratively determine the set of binding constraints at optimality, known as an *active set strategy*. This is done by solving a sequence of equality-constrained problems. See Boot [1961, 1964], Goldfarb and Idnani [1983], Powell [1985a], Theil and Panne [1961], and Panne [1974]. (Also, see Luenberger [1984] and Fletcher [1987] and Exercises 11.19 and 11.28.) Yet another approach, adopted by Houthaker [1960], is to solve a restricted problem by adding a constraint of the form $\sum x_j \leq \beta$ and successively increasing β.

One of the popular schemes for solving a quadratic program is to solve the KKT system as proposed by Barankin and Dorfman [1958] and by Markowitz [1956]. There are several methods for solving the KKT system. Wolfe [1959] developed a slight modification of the simplex method to solve the KKT system where dual feasibility is relaxed. This method was discussed briefly in Exercise 11.16. As we discussed earlier in these notes, complementary pivoting methods for solving a linear complementary problem can also be used to solve the KKT system. In Sections 11.1 and 11.2 we discussed Lemke's method for solving a quadratic program, where primal and dual feasibilities are relaxed. In Exercises 11.11, 11.16, 11.17, and 11.20 we present several alternative methods for solving the KKT system. For more details, see Cottle and Dantzig [1968], Dantzig [1963], Frank and Wolfe [1956], and Shetty [1963]. Polyak and Tret'iakov [1972] present a finite algorithm based on the ALAG penalty function approach presented in Chapter 9. Polynomial-time algorithms have also been developed for convex quadratic programs (see Ben Daya and Shetty [1988] and Ye [1989] for a survey). Primal-dual path-following algorithms in the spirit of the algorithm presented in Chapter 9 have also been proposed (see Anstreicher [1990] and Monteiro et al. [1990]). A polynomial-time algorithm, specialized to box-constrained quadratic programs is presented by Han et al. [1989]. See Ye [1990] for quadratic minimization over a sphere.

The methods discussed above deal with convex quadratic programs. Extensions to the nonconvex case have been studied by several researchers. In Exercise 11.5 the problem of finding an optimal solution is posed as the minimization of a linear objective function subject to the constraints representing a linear complementary problem. One approach for solving such problems is the use of cutting plane methods, as discussed in Balas [1975], Balas and Burdet [1973], Burdet [1977], Ritter [1966], and Tuy [1964]. Alternative approaches may be found in Cabot and Francis [1970], Mueller [1970], Mylander [1971], Taha [1973], Vanderbussche and Nemhauser [2003, 2005a,b], and Zwart [1974]. Horst and Tuy [1990] and Pardalos and Rosen [1987] survey other recent, competitive methods. Sherali [1993] discusses nonconvex quadratic programming duality. Pardalos and Vavasis [1991] show that such problems are NP-Hard, even if the Hessian has a single negative eigenvalue (for minimization problems). An algorithm for a problem that significantly generalizes quadratic programming to the case when the objective and constraint functions are general polynomials has been developed by Sherali and Tuncbilek [1992]. (See Exercise 11.27.) A specialization of this for solving nonconvex quadratic programming problems to global optimality is presented by Sherali and Tuncbilek [1995] and is discussed in Section 11.2. A further generalization of this *reformulation–linearization/convexification technique* (RLT) for solving an even wider class of *factorable programming problems* is developed by Sherali and Wang [2001]. Also, Sherali and Ganesan [2003] describe an RLT-based approach for solving more complex black-box optimization problems and apply this technique to the design of containerships. For further reading on the RLT methodology, we refer the interested reader to Sherali and Adams [1990, 1994, 1999] and the surveys in Sherali [2002] and Sherali and Desai [2004]. Also, for enhancements of the RLT methodology using *semidefinite programming* concepts, see Sherali and Fraticelli [2002].

In Section 11.3 we discuss the simplex method with a restricted basis entry rule for solving separable programming problems. Applications include economic data fitting (Bachem and Korte [1977]); electrical networks (Rockafellar [1976]); water supply system design (Collins et al. [1978] and Meyer [1980], in which problems having more than 600 constraints and 900 variables have been solved); and statistics (Teng [1978]). This approach is found in the works of Charnes and Cooper [1957], Dantzig et al. [1958], and Markowitz and Manne [1957]. For further discussion on this approach, see Miller [1963] and Wolfe [1963]. Myer [1980] discusses a novel two-segment approximation approach, and Meyer [1980] and Thakur [1978] discuss bounds on error upon early termination. In the nonconvex case, even though optimality cannot be claimed with the restricted basis entry rule, good solutions are produced. In the convex case we showed that by choosing a small grid, we can obtain a solution sufficiently close to the global optimal solution. In Section 11.3 we also discussed the grid generation scheme of Wolfe [1963]. Here grid points are not fixed beforehand but are generated as needed.

In Section 11.4 we discussed the methods of Charnes and Cooper [1962] and of Gilmore and Gomory [1963] for solving a linear fractional programming problem. The first approach makes a transformation of variables and solves an equivalent linear program. The second approach is an adaptation of the convex-simplex method. Algorithms in this category are closely related to the original work of Isbell and Marlow [1956]. Dorn [1962] presents a procedure for solving the problem that can be viewed as a generalization of the dual simplex method. For other algorithms in this general class, see Abadie and Williams [1968], Avriel et al. [1988], Bitran and Novaes [1973], Konno and Kuno [1989], Martos [1964, 1975], and Schaible [1989].

The linear fractional programming problem has been extended to the case where the objective function is the ratio of two nonlinear functions. Properties of such fractional functions are discussed in Exercises 3.11 and 3.62. Several algorithms for solving nonlinear fractional programs are developed. The interested reader may refer to Almogy and Levin [1971], Bector [1968], Dinkelbach [1967], Mangasarian [1969b], and Swarup [1965].

Geometric programming problems, discussed in Section 11.5, arise frequently in engineering applications (see Bradley and Clyne [1976], Dembo and Avriel [1978], and Duffin et al. [1967], for example). An excellent pioneering exposition appears in Duffin et al. [1967]. Exercises 11.50 through 11.55 are presented in this work, along with many other examples. We principally discuss posynomial geometric programs, following a Lagrangian duality approach (see Fletcher [1987]; see also Duffin et al. [1967] for a generalization of Theorem 11.5.2). Duffin and Peterson [1972, 1973] provide further discussions, Peterson [1976] gives a survey of approaches to a wider class of geometric programs, and Dembo [1978] and Ecker [1980] give excellent discussions on implementation and computational aspects of solving geometric programming problems. Dembo [1979] also presents details of an efficient second-order Newton-type method for solving DGP. Geometric programming problems that involve the optimization of general polynomial objective functions subject to polynomial constraints are discussed by Floudas and Visweswaran [1991], Sherali and Tuncbilek [1992], Sherali [1998], and Shor [1990]. Some test problems appear in Dembo [1976].

This page is intentionally left blank

Appendix A
Mathematical Review

In this appendix we review notation, basic definitions, and results related to vectors, matrices, and real analysis that are used throughout the text. For more details, see Bartle [1976], Berge [1963], Berge and Ghoulia-Houri [1965], Buck [1965], Cullen [1972], Flet [1966], and Rudin [1964].

A.1 Vectors and Matrices

Vectors

An n-vector $\mathbf{x}$ is an array of n scalars $x_1, x_2, ..., x_n$. Here x_j is called the jth *component*, or *element*, of the vector $\mathbf{x}$. The notation $\mathbf{x}$ represents a *column vector*, whereas the notation $\mathbf{x}^t$ represents the transposed *row vector*. Vectors are denoted by lowercase boldface letters, such as **a**, **b**, **c**, **x**, and **y**. The collection of all n-vectors forms the *n-dimensional Euclidean space*, which is denoted by R^n.

Special Vectors

The *zero vector*, denoted by 0, is a vector consisting entirely of zeros. The *sum vector* is denoted by **1** or **e** and has each component equal to 1. The ith *coordinate vector*, also referred to as the ith *unit vector*, is denoted by $\mathbf{e}_i$ and consists of zeros except for a 1 at the ith position.

Vector Addition and Multiplication by a Scalar

Let $\mathbf{x}$ and $\mathbf{y}$ be two n-vectors. The *sum* of $\mathbf{x}$ and $\mathbf{y}$ is written as the vector $\mathbf{x} + \mathbf{y}$. The jth component of the vector $\mathbf{x} + \mathbf{y}$ is $x_j + y_j$. The *product* of a vector $\mathbf{x}$ and a scalar α is denoted by $\alpha\mathbf{x}$ and is obtained by multiplying each element of $\mathbf{x}$ by α.

Linear and Affine Independence

A collection of vectors $\mathbf{x}_1, ..., \mathbf{x}_k$ in R^n is considered *linearly independent* if $\sum_{j=1}^{k} \lambda_j \mathbf{x}_j = 0$ implies that $\lambda_j = 0$ for all $j = 1, ..., k$. A collection of vectors $\mathbf{x}_0$, $\mathbf{x}_1, ..., \mathbf{x}_k$ in R^n is considered to be *affinely independent* if $(\mathbf{x}_1 - \mathbf{x}_0), ..., (\mathbf{x}_k - \mathbf{x}_0)$ are linearly independent.

Linear, Affine, and Convex Combinations and Hulls

A vector y in R^n is said to be a *linear combination* of the vectors $\mathbf{x}_1,\dots,\mathbf{x}_k$ in R^n if y can be written as $\mathbf{y} = \sum_{j=1}^{k} \lambda_j \mathbf{x}_j$ for some scalars $\lambda_1,\dots,\lambda_k$. If, in addition, $\lambda_1,\dots,\lambda_k$ are restricted to satisfy $\sum_{j=1}^{k} \lambda_j = 1$, then y is said to be an *affine combination* of $\mathbf{x}_1,\dots,\mathbf{x}_k$. Furthermore, if we also restrict $\lambda_1,\dots,\lambda_k$ to be nonnegative, then this is known as a *convex combination* of $\mathbf{x}_1,\dots,\mathbf{x}_k$. The *linear*, *affine*, or *convex hull* of a set $S \subseteq R^n$ is, respectively, the set of all linear, affine, or convex combinations of points within *S*.

Spanning Vectors

A collection of vectors $\mathbf{x}_1,\dots,\mathbf{x}_k$ in R^n, where $k \geq n$, is said to *span* R^n if any vector in R^n can be represented as a linear combination of $\mathbf{x}_1,\dots,\mathbf{x}_k$. The *cone spanned* by a collection of vectors $\mathbf{x}_1,\dots,\mathbf{x}_k$, for any $k \geq 1$, is the set of nonnegative linear combinations of these vectors.

Basis

A collection of vectors $\mathbf{x}_1,\dots,\mathbf{x}_k$ in R^n is called a *basis* of R^n if it spans R^n and if the deletion of any of the vectors prevents the remaining vectors from spanning R^n. It can be shown that $\mathbf{x}_1,\dots,\mathbf{x}_k$ form a basis of R^n if and only if $\mathbf{x}_1,\dots,\mathbf{x}_k$ are linearly independent and if, in addition, $k = n$.

Inner Product

The *inner product* of two vectors x and y in R^n is defined by $\mathbf{x}^t\mathbf{y} = \sum_{j=1}^{n} x_j y_j$. If the inner product of two vectors is equal to zero, then the two vectors are said to be *orthogonal*.

Norm of a Vector

The *norm* of a vector x in R^n is denoted by $\|\mathbf{x}\|$ and defined by $\|\mathbf{x}\| = (\mathbf{x}^t\mathbf{x})^{1/2} = (\sum_{j=1}^{n} x_j^2)^{1/2}$. This is also referred to as the ℓ_2 norm, or *Euclidean norm*.

Schwartz Inequality

Let x and y be two vectors in R^n, and let $|\mathbf{x}^t\mathbf{y}|$ denote the absolute value of $\mathbf{x}^t\mathbf{y}$. Then the following inequality, referred to as the *Schwartz inequality*, holds true:

$$|\mathbf{x}^t\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|.$$

Matrices

A *matrix* is a rectangular array of numbers. If the matrix has m rows and n columns, it is called an $m \times n$ *matrix*. Matrices are denoted by boldface capital letters, such as **A**, **B**, and **C**. The entry in row i and column j of a matrix **A** is denoted by a_{ij}, its ith row is denoted by $\mathbf{A}_i$, and its jth column is denoted by $\mathbf{a}_j$.

Special Matrices

An $m \times n$ matrix whose elements are all equal to zero is called a *zero matrix* and is denoted by **0**. A square $n \times n$ matrix is called the *identity matrix* if $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = 1$ for $i = 1,\ldots, n$. The $n \times n$ identity matrix is denoted by **I** and sometimes by $\mathbf{I}_n$ to highlight its dimension. An $n \times n$ *permutation matrix* **P** is one that has the same rows of $\mathbf{I}_n$ but which are permuted in some order. An *orthogonal matrix* **Q** having dimension $m \times n$ is one that satisfies $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}_n$ or $\mathbf{Q}\mathbf{Q}^t = \mathbf{I}_m$. In particular, if **Q** is square, $\mathbf{Q}^{-1} = \mathbf{Q}^t$. Note that a permutation matrix **P** is an orthogonal square matrix.

Addition of Matrices and Scalar Multiplication of a Matrix

Let **A** and **B** be two $m \times n$ matrices. The *sum* of **A** and **B**, denoted by **A** + **B**, is the matrix whose (i, j)th entry is $a_{ij} + b_{ij}$. The *product* of a matrix **A** by a scalar α is the matrix whose (i, j)th entry is αa_{ij}.

Matrix Multiplication

Let **A** be an $m \times n$ matrix and **B** be an $n \times p$ matrix. Then the *product* **AB** is defined to be the $m \times p$ matrix **C** whose (i, j)th entry c_{ij} is given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \quad \text{for } i = 1,\ldots, m, \text{ and } j = 1,\ldots, p.$$

Transposition

Let **A** be an $m \times n$ matrix. The *transpose* of **A**, denoted by $\mathbf{A}^t$, is the $n \times m$ matrix whose (i, j)th entry is equal to a_{ji}. A square matrix **A** is said to be *symmetric* if $\mathbf{A} = \mathbf{A}^t$. It is said to be *skew symmetric* if $\mathbf{A}^t = -\mathbf{A}$.

Partitioned Matrices

A matrix can be partitioned into submatrices. For example, the $m \times n$ matrix **A** could be partitioned as follows:

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right],$$

where $\mathbf{A}_{11}$ is $m_1 \times n_1$, $\mathbf{A}_{12}$ is $m_1 \times n_2$, $\mathbf{A}_{21}$ is $m_2 \times n_1$, $\mathbf{A}_{22}$ is $m_2 \times n_2$, $m = m_1 + m_2$, and $n = n_1 + n_2$.

Determinant of a Matrix

Let $\mathbf{A}$ be an $n \times n$ matrix. The *determinant* of $\mathbf{A}$, denoted by det[$\mathbf{A}$], is defined iteratively as follows:

$$\det[\mathbf{A}] = \sum_{i=1}^{n} a_{i1} \det[\mathbf{A}_{i1}].$$

Here $\mathbf{A}_{i1}$ is the *cofactor* of a_{i1}, defined as $(-1)^{i+1}$ times the submatrix of $\mathbf{A}$ formed by deleting the ith row and the first column, and the determinant of any scalar is the scalar itself. Similar to the use of the first column above, the determinant can be expressed in terms of any row or column.

Inverse of a Matrix

A square matrix $\mathbf{A}$ is said to be *nonsingular* if there is a matrix $\mathbf{A}^{-1}$, called the *inverse matrix*, such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The inverse of a square matrix, if it exists, is unique. Furthermore, a square matrix has an inverse if and only if its determinant is not equal to zero.

Rank of a Matrix

Let $\mathbf{A}$ be an $m \times n$ matrix. The *rank* of $\mathbf{A}$ is the maximum number of linearly independent rows or, equivalently, the maximum number of linearly independent columns of the matrix $\mathbf{A}$. If the rank of $\mathbf{A}$ is equal to $\min\{m, n\}$, $\mathbf{A}$ is said to have *full rank*.

Norm of a Matrix

Let $\mathbf{A}$ be an $n \times n$ matrix. Most commonly, the *norm* of $\mathbf{A}$, denoted by $\|\mathbf{A}\|$, is defined by

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|$$

where $\|\mathbf{A}\mathbf{x}\|$ and $\|\mathbf{x}\|$ are the usual Euclidean (ℓ_2) norms of the corresponding vectors. Hence, for any vector $\mathbf{z}$, $\|\mathbf{A}\mathbf{z}\| \le \|\mathbf{A}\|\,\|\mathbf{z}\|$. A similar use of an ℓ_p norm $\|\cdot\|_p$ induces a corresponding matrix norm $\|\mathbf{A}\|_p$. In particular, the above matrix norm, sometimes denoted $\|\mathbf{A}\|_2$, is equal to the [maximum eigenvalue of $\mathbf{A}^t\mathbf{A}]^{1/2}$. Also, the *Frobenius norm* of $\mathbf{A}$ is given by

$$\|\mathbf{A}\|_F = \left[\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right]^{1/2}$$

and is simply the ℓ_2 norm of the vector whose elements are all the elements of **A**.

Eigenvalues and Eigenvectors

Let **A** be an $n \times n$ matrix. A scalar λ and a nonzero vector **x** satisfying the equation $\mathbf{Ax} = \lambda\mathbf{x}$ are called, respectively, an *eigenvalue* and an *eigenvector* of **A**. To compute the eigenvalues of **A**, we solve the equation $\det[\mathbf{A} - \lambda\mathbf{I}] = 0$. This yields a polynomial equation in λ that can be solved for the eigenvalues of **A**. If **A** is symmetric, then it has n (possibly nondistinct) eigenvalues. The eigenvectors associated with distinct eigenvalues are necessarily orthogonal, and for any collection of some p coincident eigenvalues, there exists a collection of p orthogonal eigenvectors. Hence, given a symmetric matrix **A**, we can construct an orthogonal basis **B** for R^n, that is, a basis having orthogonal column vectors, each representing an eigenvector of **A**. Furthermore, let us assume that each column of **B** has been normalized to have a unit norm. Hence, $\mathbf{B}^t\mathbf{B} = \mathbf{I}$, so that $\mathbf{B}^{-1} = \mathbf{B}^t$. Such a matrix is said to be an *orthogonal matrix* or an *orthonormal matrix*.

Now, consider the (pure) *quadratic form* $\mathbf{x}^t\mathbf{Ax}$, where **A** is an $n \times n$ symmetric matrix. Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of **A**, let $\mathbf{\Lambda} = \text{diag}\,\{\lambda_1,\ldots,\lambda_n\}$ be a *diagonal matrix* comprised of diagonal elements $\lambda_1,\ldots,\lambda_n$ and zeros elsewhere, and let **B** be the orthogonal eigenvector matrix comprised of the orthogonal, normalized eigenvectors $\mathbf{b}_1,\ldots,\mathbf{b}_n$ as its columns. Define the linear transformation $\mathbf{x} = \mathbf{By}$ that writes any vector **x** in terms of the eigenvectors of **A**. Under this transformation, the given quadratic form becomes

$$\mathbf{x}^t\mathbf{Ax} = \mathbf{y}^t\mathbf{B}^t\mathbf{ABy} = \mathbf{y}^t\mathbf{B}^t\mathbf{\Lambda By} = \mathbf{y}^t\mathbf{\Lambda y} = \sum_{i=1}^{n}\lambda_i y_i^2.$$

This is called a *diagonalization process*.

Observe also that we have $\mathbf{AB} = \mathbf{B\Lambda}$, so that because **B** is orthogonal, we get $\mathbf{A} = \mathbf{B\Lambda B}^t = \sum_{i=1}^{n}\lambda_i\mathbf{b}_i\mathbf{b}_i^t$. This representation is called the *spectral decomposition* of **A**. For an $m \times n$ matrix **A**, a related factorization $\mathbf{A} = \mathbf{U\Sigma V}^t$, where **U** is an $m \times m$ orthogonal matrix, **V** is an $n \times n$ orthogonal matrix, and $\mathbf{\Sigma}$ is an $m \times n$ matrix having elements $\Sigma_{ij} = 0$ for $i \neq j$, and $\Sigma_{ij} \geq 0$ for $i = j$, is known as a *singular-value decomposition* (SVD) of **A**. Here, the columns of **U** and **V** are normalized eigenvectors of $\mathbf{AA}^t$ and $\mathbf{A}^t\mathbf{A}$, respectively. The Σ_{ij} values are the (absolute) square roots of the eigenvalues of $\mathbf{AA}^t$ if $m \leq n$ or of $\mathbf{A}^t\mathbf{A}$ if $m \geq n$. The number of nonzero Σ_{ij} values equals the rank of **A**.

Definite and Semidefinite Matrices

Let $\mathbf{A}$ be an $n \times n$ symmetric matrix. Here $\mathbf{A}$ is said to be *positive definite* if $\mathbf{x}^t\mathbf{A}\mathbf{x} > 0$ for all nonzero $\mathbf{x}$ in R^n and is said to be *positive semidefinite* if $\mathbf{x}^t\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$ in R^n. Similarly, if $\mathbf{x}^t\mathbf{A}\mathbf{x} < 0$ for all nonzero $\mathbf{x}$ in R^n, then $\mathbf{A}$ is called *negative definite*; and if $\mathbf{x}^t\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x}$ in R^n, then $\mathbf{A}$ is called *negative semidefinite*. A matrix that is neither positive semidefinite nor negative semidefinite is called *indefinite*. By the foregoing diagonalization process, the matrix $\mathbf{A}$ is positive definite, positive semidefinite, negative definite, and negative semidefinite if and only if its eigenvalues are positive, nonnegative, negative, and nonpositive, respectively. (Note that the superdiagonalization algorithm discussed in Chapter 3 is a more efficient method for ascertaining definiteness properties.) Also, by the definition of $\mathbf{A}$ and $\mathbf{B}$ above, if $\mathbf{A}$ is positive definite, then its *square root* $\mathbf{A}^{1/2}$ is the matrix satisfying $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$ and is given by $\mathbf{A}^{1/2} = \mathbf{B}\Lambda^{1/2}\mathbf{B}^t$.

A.2 Matrix Factorizations

Let $\mathbf{B}$ be a nonsingular $n \times n$ matrix, and consider the system of equations $\mathbf{B}\mathbf{x} = \mathbf{b}$. The solution given by $\mathbf{x} = \mathbf{B}^{-1}\mathbf{b}$ is seldom computed by finding the inverse $\mathbf{B}^{-1}$ directly. Instead, a factorization or decomposition of $\mathbf{B}$ into multiplicative components is usually employed whereby $\mathbf{B}\mathbf{x} = \mathbf{b}$ is solved in a numerically stable fashion, often through the solution of triangular systems via back-substitution. This becomes particularly relevant in ill-conditioned situations when $\mathbf{B}$ is nearly singular or when we wish to verify positive definiteness of $\mathbf{B}$ as in quasi-Newton or Levenberg-Marquardt methods. Several useful factorizations are discussed below. For more details, including schemes for updating such factors in the context of iterative methods, we refer the reader to Bartels et al. [1970], Bazaraa et al. [2005], Dennis and Schnabel [1983], Dongarra et al. [1979], Gill et al. [1974, 1976], Golub and Van Loan [1983/1989], Murty [1983], and Stewart [1973], along with the many accompanying references cited therein. Standard software such as LINPACK, MATLAB, and the Harwell Library routines are also available to perform these factorizations efficiently.

LU and PLU Factorization for a Basis B

In the LU factorization, we reduce $\mathbf{B}$ to an upper triangular form $\mathbf{U}$ through a series of permutations and Gaussian pivot operations. At the ith stage of this process, having reduced $\mathbf{B}$ to $\mathbf{B}^{(i-1)}$, say, which is upper triangular in columns $1,\ldots, i-1$ (where $\mathbf{B}^0 \equiv \mathbf{B}$), we first premultiply $\mathbf{B}^{(i-1)}$ by a *permutation matrix* $\mathbf{P}_i$ to exchange row i with that row in $\{i, i+1,\ldots, n\}$ of $\mathbf{B}^{(i-1)}$ that has the largest absolute-valued element in column i. This is done to ensure that the (i, i)th element of $\mathbf{P}_i\mathbf{B}^{(i-1)}$ is significantly nonzero. Using this as a pivot element, we perform row operations to zero out the elements in rows $i + 1,..., n$ of column i. This *triangularization* can be represented as a premultiplication with a suitable *Gaussian pivot matrix* $\mathbf{G}_i$, which is a *unit lower triangular matrix*, having ones on the diagonal and suitable possibly nonzero elements in rows $i + 1,..., n$ of column i. This gives $\mathbf{B}^{(i)} = (\mathbf{G}_i\mathbf{P}_i)\mathbf{B}^{(i-1)}$. Hence, we get, after some $r \le (n-1)$ such operations,

$$(\mathbf{G}_r\mathbf{P}_r)\cdots(\mathbf{G}_2\mathbf{P}_2)(\mathbf{G}_1\mathbf{P}_1)\mathbf{B} = \mathbf{U}. \tag{A.1}$$

The system $\mathbf{Bx} = \mathbf{b}$ can now be solved by computing $\bar{\mathbf{b}} = (\mathbf{G}_r\mathbf{P}_r)\cdots(\mathbf{G}_1\mathbf{P}_1)\mathbf{b}$ and then solving the triangular system $\mathbf{Ux} = \bar{\mathbf{b}}$ by back-substitution. If no permutations are performed, $\mathbf{G}_r\cdots\mathbf{G}_1$ is lower triangular, and denoting its (lower triangular) inverse as $\mathbf{L}$, we have the factored form $\mathbf{B} = \mathbf{LU}$ for $\mathbf{B}$, hence its name. Also, if $\mathbf{P}^t$ is a permutation matrix that is used to *a priori* rearrange the rows of $\mathbf{B}$ and we then apply the Gaussian triangularization operation to derive $\mathbf{L}^{-1}\mathbf{P}^t\mathbf{B} = \mathbf{U}$, we can write $\mathbf{B} = (\mathbf{P}^t)^{-1}\mathbf{LU} = \mathbf{PLU}$, noting that $\mathbf{P}^t = \mathbf{P}^{-1}$. Hence, this factorization is sometimes called a *PLU decomposition*. If $\mathbf{B}$ is sparse, $\mathbf{P}^t$ can be used to make $\mathbf{P}^t\mathbf{B}$ nearly upper triangular (assuming that the columns of $\mathbf{B}$ have been appropriately permuted) and then only a few and sparse Gaussian pivot operations will be required to obtain $\mathbf{U}$. This method is therefore very well suited for sparse matrices.

QR and *QRP* Factorization for a Basis B

This factorization is most suitable and is used frequently for solving *dense equation systems*. Here the matrix $\mathbf{B}$ is reduced to an upper triangular form $\mathbf{R}$ by premultiplying it with a sequence of square, symmetric orthogonal matrices $\mathbf{Q}_i$. Given $\mathbf{B}^{(i-1)} \equiv \mathbf{Q}_{i-1}\cdots\mathbf{Q}_1\mathbf{B}$ that is upper triangular in columns $1,..., i-1$ (where $\mathbf{B}^{(0)} = \mathbf{B}$), we construct a matrix $\mathbf{Q}_i$ so that $\mathbf{Q}_i\mathbf{B}^{(i-1)} = \mathbf{B}^{(i)}$ is upper triangular in column i as well, while columns $1,..., i-1$ remain unaffected. The matrix $\mathbf{Q}_i$ is a square, symmetric orthogonal matrix of the form $\mathbf{Q}_i \equiv \mathbf{I} - \gamma_i\mathbf{q}_i\mathbf{q}_i^t$, where $\mathbf{q}_i = (0,...,0, q_{ii},..., q_{ni})^t$ and $\gamma_i \in R^1$ are suitably chosen to perform the foregoing operation. Such a matrix $\mathbf{Q}_i$ is called a *Householder transformation matrix*. If the elements in rows $i,..., n$ of column i of $\mathbf{B}^{(i-1)}$ are denoted by $(\alpha_i,...,\alpha_n)^t$, then we have $q_{ii} = \alpha_i + \theta_i$, $q_{ji} = \alpha_j$ for $j = i + 1,..., n$, $\gamma_i = 1/\theta_i q_{ii}$, where $\theta_i = \text{sign}(\alpha_i)[\alpha_i^2 + \cdots + \alpha_n^2]^{1/2}$, and where $\text{sign}(\alpha_i) = 1$ if $\alpha_i > 0$ and -1 otherwise. Defining $\mathbf{Q} = \mathbf{Q}_{n-1}\cdots\mathbf{Q}_1$, we see that $\mathbf{Q}$ is also a symmetric orthogonal matrix and that $\mathbf{QB} = \mathbf{R}$, or that $\mathbf{B} = \mathbf{QR}$, since $\mathbf{Q} = \mathbf{Q}^t = \mathbf{Q}^{-1}$; that is, $\mathbf{Q}$ is an *involutory matrix*.

Now, to solve $\mathbf{Bx} = \mathbf{b}$, we equivalently solve $\mathbf{QRx} = \mathbf{b}$ or $\mathbf{Rx} = \mathbf{Qb}$ by finding $\bar{\mathbf{b}} = \mathbf{Qb}$ first and then solving the upper triangular system $\mathbf{Rx} = \bar{\mathbf{b}}$ via back-substitution. Note that since $\|\mathbf{Qv}\| = \|\mathbf{v}\|$ for any vector v, we have $\|\mathbf{R}\| = \|\mathbf{QR}\| = \|\mathbf{B}\|$, so that $\mathbf{R}$ preserves the relative magnitudes of the elements in $\mathbf{B}$, maintaining stability. This is its principal advantage.

Also, a permutation matrix $\mathbf{P}$ is sometimes used to postmultiply $\mathbf{B}^{(i-1)}$ before applying $\mathbf{Q}_i$ to it, so as to move a column that has the largest value of the sum of squares below row $i-1$ into the ith column position (see the computation of θ_i above). Since the product of permutation matrices is also a permutation matrix, and since a permutation matrix is orthogonal, this leads to the decomposition $\mathbf{B} = \mathbf{QRP}$ via the operation sequence $\mathbf{Q}_{n-1}\cdots\mathbf{Q}_1\mathbf{BP}_1\mathbf{P}_2\cdots\mathbf{P}_{n-1} = \mathbf{R}$.

Cholesky Factorization LL^t and LDL^t for Symmetric, Positive Definite Matrices B

The Cholesky factorization of a symmetric, positive definite matrix $\mathbf{B}$ represents this matrix as $\mathbf{B} = \mathbf{LL}^t$, where $\mathbf{L}$ is a lower triangular matrix of the form

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & & & & \\ \ell_{21} & \ell_{22} & & \mathbf{0} & \\ \ell_{31} & \ell_{32} & \ell_{33} & & \\ \vdots & \vdots & & \ddots & \\ \ell_{n1} & \ell_{n2} & & & \ell_{nn} \end{bmatrix},$$

so that

$$\mathbf{LL}^t = \begin{bmatrix} \ell_{11}^2 & & & & \text{(symmetric)} \\ \ell_{21}\ell_{11} & (\ell_{21}^2+\ell_{22}^2) & & & \\ \ell_{31}\ell_{11} & (\ell_{21}\ell_{31}+\ell_{22}\ell_{32}) & (\ell_{31}^2+\ell_{32}^2+\ell_{33}^2) & & \\ \vdots & \vdots & \vdots & & \ddots \\ \ell_{n1}\ell_{11} & (\ell_{21}\ell_{n1}+\ell_{22}\ell_{n2}) & (\ell_{31}\ell_{n1}+\ell_{32}\ell_{n2}+\ell_{33}\ell_{n3}) & \cdots & (\ell_{n1}^2+\cdots+\ell_{nn}^2) \end{bmatrix}.$$

By equating the elements of $\mathbf{B}$ directly to those in $\mathbf{LL}^t$, we obtain the system of equations

$$\begin{array}{lllll} \ell_{11}^2 = b_{11}, & \ell_{21}\ell_{11} = b_{21}, & \ell_{31}\ell_{11} = b_{31}, & \ldots, & \ell_{n1}\ell_{11} = b_{n1} \\ & \ell_{21}^2+\ell_{22}^2 = b_{22}, & \ell_{21}\ell_{31}+\ell_{22}\ell_{32} = b_{32}, & \ldots, & \ell_{21}\ell_{n1}+\ell_{22}\ell_{n2} = b_{n2} \\ & & \ell_{31}^2+\ell_{32}^2+\ell_{33}^2 = b_{33}, & \ldots, & \ell_{31}\ell_{n1}+\ell_{32}\ell_{n2}+\ell_{33}\ell_{n3} = b_{n3} \\ & & & \ddots & \vdots \\ & & & & \ell_{n1}^2+\cdots+\ell_{nn}^2 = b_{nn}. \end{array}$$

These equations can be used sequentially to compute the unknowns ℓ_{ij} in the order $\ell_{11}, \ell_{21},\ldots,\ell_{n1}, \ell_{22}, \ell_{32},\ldots,\ell_{n2}, \ell_{33},\ldots,\ell_{n3},\ldots,\ell_{nn}$, by using the equation for b_{ij} to compute ℓ_{ij} for $j = 1,..., n$, $i = j,..., n$. Note that these equations are well-defined for a symmetric, positive definite matrix **B** and that $\mathbf{LL}^t$ is positive definite if and only if $\ell_{ii} > 0$ for all $i = 1,..., n$.

The equation system $\mathbf{Bx} = \mathbf{b}$ can now be solved via $\mathbf{L}(\mathbf{L}^t\mathbf{x}) = \mathbf{b}$ through the solution of two triangular systems of equations. We first find **y** to satisfy $\mathbf{Ly} = \mathbf{b}$ and then compute **x** via the system $\mathbf{L}^t\mathbf{x} = \mathbf{y}$.

Sometimes the Cholesky factorization is represented as $\mathbf{B} = \mathbf{LDL}^t$, where **L** is a lower triangular matrix (usually having ones along its diagonal) and **D** is a diagonal matrix, both having positive diagonal entries. Writing $\mathbf{B} = \mathbf{LDL}^t = (\mathbf{LD}^{1/2})(\mathbf{LD}^{1/2})^t \equiv \mathbf{L}'\mathbf{L}'^t$, we see that the two representations are related equivalently. The advantage of the representation $\mathbf{LDL}^t$ is that **D** can be used to avoid the square root operation associated with the diagonal system of equations, and this improves the accuracy of computations. (For example, the diagonal components of **L** can be made unity.)

Also, if **B** is a general basis matrix, then since $\mathbf{BB}^t$ is symmetric and positive definite, it has a Cholesky factorization $\mathbf{BB}^t = \mathbf{LL}^t$. In such a case, **L** is referred to as the *Cholesky factor associated with* **B**. Note that we can determine **L** in this case by finding the *QR decomposition* for $\mathbf{B}^t$ so that $\mathbf{BB}^t = \mathbf{R}^t\mathbf{Q}^t\mathbf{QR} = \mathbf{R}^t\mathbf{R}$, and therefore, $\mathbf{L} \equiv \mathbf{R}^t$. Whenever this is done, note that the matrix **Q** or its components $\mathbf{Q}_i$ need not be stored, since we are only interested in the resulting upper triangular matrix **R**.

A.3 Sets and Sequences

A *set* is a collection of elements or objects. A set may be specified by listing its elements or by specifying the properties that the elements must satisfy. For example, the set $S = \{1, 2, 3, 4\}$ can be represented alternatively as $S = \{x : 1 \le x \le 4, x \text{ integer}\}$. If **x** is a member of S, we write $\mathbf{x} \in S$, and if **x** is not a member of S, we write $\mathbf{x} \notin S$. Sets are denoted by capital letters, such as S, X, and A. The *empty set*, denoted by $\varnothing$, has no elements.

Unions, Intersections, and Subsets

Given two sets, S_1 and S_2, the set consisting of elements that belong to either S_1 or S_2, or both, is called the *union* of S_1 and S_2 and is denoted by $S_1 \cup S_2$. The elements belonging to both S_1 and S_2 form the *intersection* of S_1 and S_2, denoted $S_1 \cap S_2$. If S_1 is a *subset* of S_2, that is, if each element of S_1 is also an element of S_2, we write $S_1 \subseteq S_2$ or $S_2 \supseteq S_1$. Thus, we write $S \subseteq R^n$ to denote that all elements in S are points in R^n. A *strict containment* $S_1 \subseteq S_2$, $S_1 \neq S_2$, is denoted by $S_1 \subset S_2$.

Closed and Open Intervals

Let a and b be two real numbers. The *closed interval* $[a, b]$ denotes all real numbers satisfying $a \leq x \leq b$. Real numbers satisfying $a \leq x < b$ are represented by $[a, b)$, while those satisfying $a < x \leq b$ are denoted by $(a, b]$. Finally, the set of points x with $a < x < b$ is represented by the *open interval* (a, b).

Greatest Lower Bound and Least Upper Bound

Let S be a set of real numbers. Then the ***greatest lower bound***, or the ***infimum***, of S is the largest possible scalar α satisfying $\alpha \leq x$ for each $x \in S$. The infimum is denoted by $\inf \{x : x \in S\}$. The *least upper bound*, or the ***supremum***, of S is the smallest possible scalar α satisfying $\alpha \geq x$ for each $x \in S$. The supremum is denoted by $\sup \{x : x \in S\}$.

Neighborhoods

Given a point $\mathbf{x} \in R^n$ and an $\varepsilon > 0$, the *ball* $N_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\}$ is called an *ε-neighborhood* of $\mathbf{x}$. The inequality in the definition of $N_\varepsilon(\mathbf{x})$ is sometimes replaced by a strict inequality.

Interior Points and Open Sets

Let S be a subset of R^n, and let $\mathbf{x} \in S$. Then $\mathbf{x}$ is called an *interior point* of S if there is an ε-neighborhood of $\mathbf{x}$ that is contained in S, that is, if there exists an $\varepsilon > 0$ such that $\|\mathbf{y} - \mathbf{x}\| \leq \varepsilon$ implies that $\mathbf{y} \in S$. The set of all such points is called the *interior* of S and is denoted by int S. Furthermore, S is called *open* if S = int S.

Relative Interior

Let $S \subset R^n$, and let aff(S) denote the *affine hull* of S. Although int(S) = $\varnothing$, the interior of S as viewed in the space of its affine hull may be nonempty. This is called the *relative interior* of S and is denoted by relint(S). Specifically, relint(S) = $\{\mathbf{x} \in S : N_\varepsilon(\mathbf{x}) \cap \text{aff}(S) \subset S \text{ for some } \varepsilon > 0\}$. Note that if $S_1 \subseteq S_2$, relint(S_1) is not necessarily contained within relint(S_2), although int(S_1) $\subseteq$ int(S_2). For example, if $S_1 = \{\mathbf{x} : \alpha^t \mathbf{x} = \beta\}$, $\alpha \neq 0$ and $S_2 = \{\mathbf{x} : \alpha^t \mathbf{x} \leq \beta\}$, $S_1 \subseteq S_2$, int(S_1) = $\varnothing \subseteq$ int(S_2) = $\{\mathbf{x} : \alpha^t \mathbf{x} < \beta\}$, but relint($S_1$) = $S_1 \not\subseteq$ relint(S_2) = int(S_2).

Bounded Sets

A set $S \subset R^n$ is said to be *bounded* if it can be contained within a ball of finite radius.

Closure Points and Closed Sets

Let S be a subset of R^n. The *closure* of S, denoted cl S, is the set of all points that are arbitrarily close to S. In particular, $\mathbf{x} \in \text{cl } S$ if for each $\varepsilon > 0$, $S \cap N_\varepsilon(\mathbf{x}) \neq \varnothing$, where $N_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y}-\mathbf{x}\| \leq \varepsilon\}$. The set S is said to be *closed* if $S = \text{cl } S$.

Boundary Points

Let S be a subset of R^n. Then $\mathbf{x}$ is called a *boundary point* of S if for each $\varepsilon > 0$, $N_\varepsilon(\mathbf{x})$ contains a point in S and a point not in S, where $N_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y}-\mathbf{x}\| \leq \varepsilon\}$. The set of all boundary points is called the *boundary* of S and is denoted by ∂S.

Sequences and Subsequences

A *sequence* of vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$, is said to *converge* to the *limit point* $\bar{\mathbf{x}}$ if $\|\mathbf{x}_k - \bar{\mathbf{x}}\| \to 0$ as $k \to \infty$; that is, if for any given $\varepsilon > 0$, there is a positive integer N such that $\|\mathbf{x}_k - \bar{\mathbf{x}}\| < \varepsilon$ for all $k \geq N$. The sequence is usually denoted by $\{\mathbf{x}_k\}$, and the limit point $\bar{\mathbf{x}}$ is represented by either $\mathbf{x}_k \to \bar{\mathbf{x}}$ as $k \to \infty$ or by $\lim_{k\to\infty} \mathbf{x}_k = \bar{\mathbf{x}}$. Any converging sequence has a unique limit point.

By deleting certain elements of a sequence $\{\mathbf{x}_k\}$, we obtain a *subsequence*. A subsequence is usually denoted as $\{\mathbf{x}_k\}_{\mathscr{K}}$ where $\mathscr{K}$ is a subset of all positive integers. To illustrate, let $\mathscr{K}$ be the set of all even positive integers. Then $\{\mathbf{x}_k\}_{\mathscr{K}}$ denotes the subsequence $\{\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_6, \ldots\}$.

Given a subsequence $\{\mathbf{x}_k\}_{\mathscr{K}}$, the notation $\{\mathbf{x}_{k+1}\}_{\mathscr{K}}$ denotes the subsequence obtained by adding 1 to the indices of all elements in the subsequence $\{\mathbf{x}_k\}_{\mathscr{K}}$. To illustrate, if $\mathscr{K} = \{3, 5, 10, 15, \ldots,\}$, then $\{\mathbf{x}_{k+1}\}_{\mathscr{K}}$ denotes the subsequence $\{\mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_{11}, \mathbf{x}_{16}, \ldots\}$.

A sequence $\{\mathbf{x}_k\}$ is called a *Cauchy sequence* if for any given $\varepsilon > 0$, there is a positive integer N such that $\|\mathbf{x}_k - \mathbf{x}_m\| < \varepsilon$ for all $k, m \geq N$. A sequence in R^n has a limit if and only if it is Cauchy.

Let $\{x_n\}$ be a bounded sequence in R. The *limit superior* of $\{x_n\}$, denoted limsup(x_n) or $\overline{\lim}(x_n)$, equals the infimum of all numbers $q \in R$ for which at most a finite number of the elements of $\{x_n\}$ (strictly) exceed q. Similarly, the *limit inferior* of $\{x_n\}$ is given by $\liminf(x_n) \equiv \underline{\lim}(x_n) \equiv \sup\{q$: at most a finite number of elements of $\{x_n\}$ are (strictly) less than $q\}$. A bounded sequence always has a unique $\overline{\lim}$ and $\underline{\lim}$.

Compact Sets

A set S in R^n is said to be *compact* if it is closed and bounded. For every sequence $\{\mathbf{x}_k\}$ in a compact set S, there is a convergent subsequence with a limit in S.

A.4 Functions

A *real-valued function* f defined on a subset S of R^n associates with each point $\mathbf{x}$ in S a real number $f(\mathbf{x})$. The notation $f: S \to R$ denotes that the domain of f is S and that the range is a subset of the real numbers. If f is defined everywhere on R^n or if the domain is not important, the notation $f: R^n \to R$ is used. A collection of real-valued functions $f_1,\ldots,f_m$ can be viewed as a single *vector function* $\mathbf{f}$ whose jth component is f_j.

Continuous Functions

A function $f: S \to R$ is said to be *continuous at* $\overline{\mathbf{x}} \in S$ if for any given $\varepsilon > 0$, there is a $\delta > 0$ such that $\mathbf{x} \in S$ and $\|\mathbf{x}-\overline{\mathbf{x}}\| < \delta$ imply that $|f(\mathbf{x})-f(\overline{\mathbf{x}})| < \varepsilon$. Equivalently, f is continuous at $\overline{\mathbf{x}} \in S$, if for any sequence $\{\mathbf{x}_n\} \to \overline{\mathbf{x}}$ such that $\{f(\mathbf{x}_n)\} \to \overline{f}$, we have that $f(\overline{\mathbf{x}}) = \overline{f}$ as well. A vector-valued function is said to be continuous at $\overline{\mathbf{x}}$ if each of its components is continuous at $\overline{\mathbf{x}}$.

Upper and Lower Semicontinuity

Let S be a nonempty set in R^n. A function $f: S \to R$ is said to be *upper semicontinuous* at $\overline{\mathbf{x}} \in S$ if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\mathbf{x} \in S$ and $\|\mathbf{x}-\overline{\mathbf{x}}\| < \delta$ imply that $f(\mathbf{x}) - f(\overline{\mathbf{x}}) < \varepsilon$. Similarly, a function $f: R^n \to R$ is called *lower semicontinuous at* $\overline{\mathbf{x}} \in S$ if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\mathbf{x} \in S$ and $\|\mathbf{x}-\overline{\mathbf{x}}\| < \delta$ imply that $f(\mathbf{x}) - f(\overline{\mathbf{x}}) > -\varepsilon$. Equivalently, then f is *upper semicontinuous at* $\overline{\mathbf{x}} \in S$, if, for any sequence $\{\mathbf{x}_n\} \to \overline{\mathbf{x}}$ such that $\{f(\mathbf{x}_n)\} \to \overline{f}$, we have $f(\overline{\mathbf{x}}) \geq \overline{f}$. Similarly, if $f(\overline{\mathbf{x}}) \leq \overline{f}$ for any such sequence, then f is said to be *lower semicontinuous at* $\overline{\mathbf{x}}$. Hence, a function is *continuous at* $\overline{\mathbf{x}}$ if and only if it is both upper and lower semicontinuous at $\overline{\mathbf{x}}$. A vector-valued function is called upper or lower semicontinuous if each of its components is upper or lower semicontinuous, respectively.

Minima and Maxima of Semicontinuous Functions

Let S be a nonempty compact set in R^n and suppose that $f: R^n \to R$. If f is lower semicontinuous, then it assumes a minimum over S; that is, there exists an $\overline{\mathbf{x}} \in S$

such that $f(\mathbf{x}) \le f(\overline{\mathbf{x}})$ for each $\mathbf{x} \in S$. Similarly, if f is upper semicontinuous, then it assumes a maximum over S. Since a continuous function is both lower and upper semicontinuous, it achieves both a minimum and a maximum over any nonempty compact set.

Differentiable Functions

Let S be a nonempty set in R^n, $\overline{\mathbf{x}} \in \text{int } S$ and let $f: S \to R$. Then f is said to be *differentiable at* $\overline{\mathbf{x}}$ if there is a vector $\nabla f(\overline{\mathbf{x}})$ in R^n called the *gradient* of f at $\overline{\mathbf{x}}$ and a function β satisfying $\beta(\overline{\mathbf{x}}; \mathbf{x}) \to 0$ as $\mathbf{x} \to \overline{\mathbf{x}}$ such that

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + \nabla f(\overline{\mathbf{x}})^t(\mathbf{x} - \overline{\mathbf{x}}) + \|\mathbf{x} - \overline{\mathbf{x}}\| \, \beta(\overline{\mathbf{x}}; \mathbf{x}) \qquad \text{for each } \mathbf{x} \in S.$$

The gradient vector consists of the partial derivatives, that is,

$$\nabla f(\overline{\mathbf{x}})^t = \left(\frac{\partial f(\overline{\mathbf{x}})}{\partial x_1}, \frac{\partial f(\overline{\mathbf{x}})}{\partial x_2}, \dots, \frac{\partial f(\overline{\mathbf{x}})}{\partial x_n} \right).$$

Furthermore, f is called *twice differentiable at* $\overline{\mathbf{x}}$ if, in addition to the gradient vector, there exist an $n \times n$ symmetric matrix $\mathbf{H}(\overline{\mathbf{x}})$, called the *Hessian matrix* of f at $\overline{\mathbf{x}}$, and a function β satisfying $\beta(\overline{\mathbf{x}}; \mathbf{x}) \to 0$ as $\mathbf{x} \to \overline{\mathbf{x}}$ such that

$$f(\mathbf{x}) = f(\overline{\mathbf{x}}) + \nabla f(\overline{\mathbf{x}})^t(\mathbf{x} - \overline{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \overline{\mathbf{x}})^t \mathbf{H}(\overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}}) + \|\mathbf{x} - \overline{\mathbf{x}}\|^2 \beta(\overline{\mathbf{x}}; \mathbf{x})$$

$$\text{for each } \mathbf{x} \in S.$$

The element in row i and column j of the Hessian matrix is the second partial $\partial^2 f(\overline{\mathbf{x}})/\partial x_i \, \partial x_j$.

A vector-valued function is differentiable if each of its components is differentiable and is twice differentiable if each of its components is twice differentiable.

In particular, for a differentiable vector function $\mathbf{h}: R^n \to R^\ell$ where $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \dots, h_\ell(\mathbf{x}))^t$, the *Jacobian* of $\mathbf{h}$, denoted by the gradient notation $\nabla \mathbf{h}(\mathbf{x})$, is given by the $\ell \times n$ matrix

$$\nabla \mathbf{h}(\mathbf{x}) = \begin{bmatrix} \nabla h_1(\mathbf{x})^t \\ \vdots \\ \nabla h_\ell(\mathbf{x})^t \end{bmatrix}_{\ell \times n},$$

whose rows correspond to the transpose of the gradients of $h_1, \dots, h_\ell$, respectively.

Mean Value Theorem

Let S be a nonempty open convex set in R^n, and let f: $S \to R$ be differentiable. The mean value theorem can be stated as follows. For every $\mathbf{x}_1$ and $\mathbf{x}_2$ in S, we must have

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\mathbf{x})^t (\mathbf{x}_2 - \mathbf{x}_1),$$

where $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\lambda \in (0, 1)$.

Taylor's Theorem

Let S be a nonempty open convex set in R^n, and let f: $S \to R$ be twice differentiable. The second-order form of *Taylor's theorem* can be stated as follows. For every $\mathbf{x}_1$ and $\mathbf{x}_2$ in S, we must have

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1) + \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)^t \mathbf{H}(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1),$$

where $\mathbf{H}(\mathbf{x})$ is the Hessian of f at $\mathbf{x}$, and where $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\lambda \in$ (0, 1).

Appendix B

Summary of Convexity, Optimality Conditions, and Duality

This appendix gives a summary of the relevant results from Chapters 2 through 6 on convexity, optimality conditions, and duality. *It is intended to provide the minimal background needed for an adequate coverage of Chapters 8 through 11, excluding convergence analysis.*

B.1 Convex Sets

A set S in R^n is said to be *convex* if for each $\mathbf{x}_1, \mathbf{x}_2 \in S$, the *line segment* $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for $\lambda \in [0, 1]$ belongs to S. Points of the form $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for $\lambda \in [0, 1]$ are called *convex combinations* of $\mathbf{x}_1$ and $\mathbf{x}_2$. Figure B.1 illustrates an example of a convex set and an example of a nonconvex set.

We present below some examples of convex sets frequently encountered in mathematical programming.

1. ***Hyperplane***: $S = \{\mathbf{x} : \mathbf{p}^t\mathbf{x} = \alpha\}$, where $\mathbf{p}$ is a nonzero vector in R^n, called the *normal* to the hyperplane, and α is a scalar.
2. ***Half-space***: $S = \{\mathbf{x} : \mathbf{p}^t\mathbf{x} \le \alpha\}$, where $\mathbf{p}$ is a nonzero vector in R^n and α is a scalar.
3. ***Open half-space***: $S = \{\mathbf{x} : \mathbf{p}^t\mathbf{x} < \alpha\}$, where $\mathbf{p}$ is a nonzero vector in R^n and α is a scalar.
4. ***Polyhedral set***: $S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \le \mathbf{b}\}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an m-vector .
5. ***Polyhedral cone***: $S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \le \mathbf{0}\}$, where $\mathbf{A}$ is an $m \times n$ matrix.
6. ***Cone spanned by a finite number of vectors***: $S = \{\mathbf{x} : \mathbf{x} = \sum_{j=1}^{m} \lambda_j \mathbf{a}_j,\ \lambda_j \ge 0 \text{ for } j = 1,\dots, m\}$, where $\mathbf{a}_1,\dots,\mathbf{a}_m$ are given vectors in R^n.
7. ***Neighborhood***: $S = \{\mathbf{x} : \|\mathbf{x} - \bar{\mathbf{x}}\| \le \varepsilon\}$, where $\bar{\mathbf{x}}$ is a fixed vector in R^n and $\varepsilon > 0$.

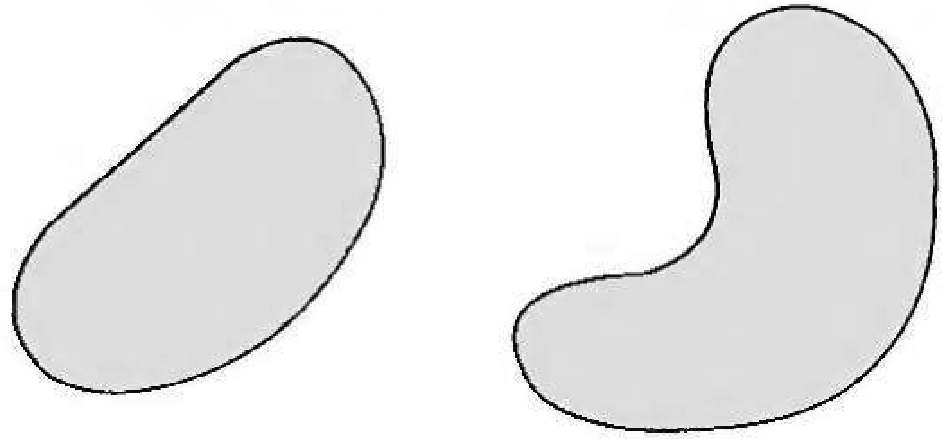


Figure B.1 Convexity.

Given two nonempty convex sets S_1 and S_2 in R^n such that $S_1 \cap S_2 = \varnothing$, there exists a hyperplane $H = \{\mathbf{x} : \mathbf{p}^t\mathbf{x} = \alpha\}$ that separates them; that is,

$$\mathbf{p}^t\mathbf{x} \le \alpha \text{ for all } \mathbf{x} \in S_1 \qquad \text{and} \qquad \mathbf{p}^t\mathbf{x} \ge \alpha \text{ for all } \mathbf{x} \in S_2.$$

Here H is called a *separating hyperplane* whose normal is the nonzero vector $\mathbf{p}$.

Closely related to the above concept is the notion of a *supporting hyperplane*. Let S be a nonempty convex set in R^n, and let $\bar{\mathbf{x}}$ be a boundary point. Then there exists a hyperplane $H = \{\mathbf{x} : \mathbf{p}^t\mathbf{x} = \alpha\}$ that supports S at $\bar{\mathbf{x}}$; that is,

$$\mathbf{p}^t\bar{\mathbf{x}} = \alpha \qquad \text{and} \qquad \mathbf{p}^t\mathbf{x} \le \alpha \text{ for all } \mathbf{x} \in S.$$

In Figure B.2 we illustrate the concepts of separating and supporting hyperplanes.

The following two theorems are used in proving optimality conditions and duality relationships and in developing termination criteria for algorithms.

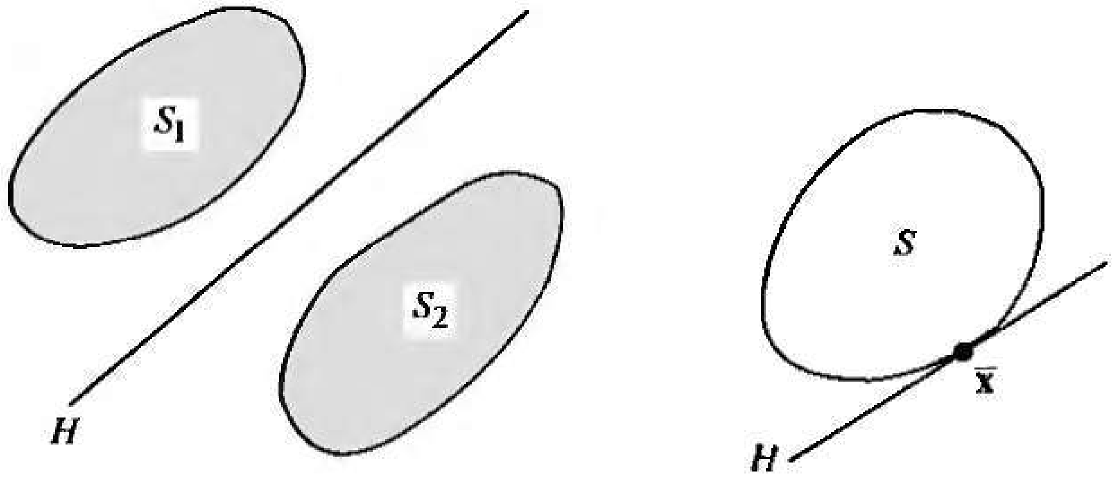


Figure B.2 Separating and supporting hyperplanes.

Farkas's Theorem

Let **A** be an $m \times n$ matrix and let **c** be an n-vector. Then exactly one of the following two systems has a solution:

System 1 $\quad \mathbf{Ax} \le \mathbf{0}, \mathbf{c}^t\mathbf{x} > 0 \qquad$ for some $\mathbf{x} \in R^n$.

System 2 $\quad \mathbf{A}^t\mathbf{y} = \mathbf{c}, \mathbf{y} \ge \mathbf{0} \qquad$ for some $\mathbf{y} \in R^m$.

Gordan's Theorem

Let **A** be an $m \times n$ matrix. Then exactly one of the following systems has a solution.

System 1 $\quad \mathbf{Ax} < \mathbf{0} \qquad$ for some $\mathbf{x} \in R^n$.

System 2 $\quad \mathbf{A}^t\mathbf{y} = \mathbf{0}, \mathbf{y} \ge \mathbf{0} \qquad$ for some nonzero $\mathbf{y} \in R^m$.

An important concept in convexity is that of an extreme point. Let S be a non-empty convex set in R^n. A vector $\mathbf{x} \in S$ is called an *extreme point* of S if $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ with $\mathbf{x}_1, \mathbf{x}_2 \in S$, and $\lambda \in (0, 1)$ implies that $\mathbf{x} = \mathbf{x}_1 = \mathbf{x}_2$. In other words, **x** is an extreme point if it cannot be represented as a strict convex combination of two distinct points in S. In particular, for the set $S = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \ge \mathbf{0}\}$, where **A** is an $m \times n$ matrix of rank m and **b** is an m-vector, **x** is an *extreme point* of S if and only if the following conditions hold true. The matrix **A** can be decomposed into [**B**, **N**], where **B** is an $m \times m$ invertible matrix and $\mathbf{x}^t = (\mathbf{x}_B^t, \mathbf{x}_N^t)$, where $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} \ge \mathbf{0}$ and $\mathbf{x}_N = \mathbf{0}$.

Another concept that is used in the case of an unbounded convex set is that of a direction of the set. Specifically, if S is an unbounded closed convex set, a vector **d** is a *direction* of S if $\mathbf{x} + \lambda\mathbf{d} \in S$ for each $\lambda \ge 0$ and for each $\mathbf{x} \in S$.

B.2 Convex Functions and Extensions

Let S be a nonempty convex set in R^n. The function f: $S \to R$ is said to be *convex* on S if

$$f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] \le \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)$$

for each $\mathbf{x}_1, \mathbf{x}_2 \in S$ and for each $\lambda \in [0, 1]$. The function f is said to be *strictly convex* on S if the above inequality holds as a strict inequality for each distinct $\mathbf{x}_1, \mathbf{x}_2 \in S$ and for each $\lambda \in (0, 1)$. The function f is said to be *concave* (*strictly concave*) if $-f$ is convex (strictly convex). Figure B.3 shows some examples of convex and concave functions.

Following are some examples of convex functions. By taking the negatives of these functions, we get some examples of concave functions.

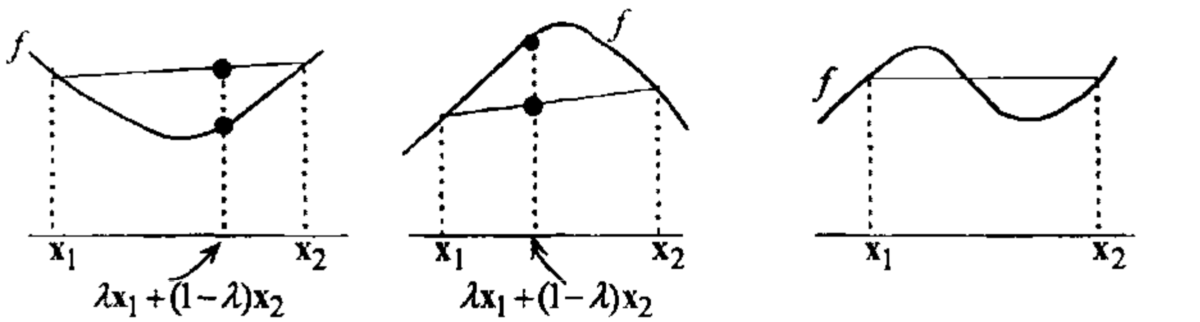


Figure B.3 Convex and concave functions.

1. $f(x) = 3x + 4$.
2. $f(x) = |x|$.
3. $f(x) = x^2 - 2x$.
4. $f(x) = -x^{1/2}$ for $x \geq 0$.
5. $f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1x_2$.
6. $f(x_1, x_2, x_3) = x_1^4 + 2x_2^2 + 3x_3^2 - 4x_1 - 4x_2x_3$.

In many cases, the assumption of convexity of a function can be relaxed to the weaker notions of quasiconvex and pseudoconvex functions.

Let S be a nonempty convex set in R^n. The function $f: S \to R$ is said to be *quasiconvex* on S if for each $\mathbf{x}_1$, $\mathbf{x}_2 \in S$, the following inequality holds true:

$$f[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] \leq \max\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \qquad \text{for each } \lambda \in (0, 1).$$

The function f is said to be *strictly quasiconvex* on S if the above inequality holds as a strict inequality, provided that $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$. The function f is said to be *strongly quasiconvex* on S if the above inequality holds as a strict inequality for $\mathbf{x}_1 \neq \mathbf{x}_2$.

Let S be a nonempty open convex set in R^n. The function $f: S \to R$ is said to be *pseudoconvex* if for each $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ with $\nabla f(\mathbf{x}_1)^t(\mathbf{x}_2 - \mathbf{x}_1) \geq 0$, we have $f(\mathbf{x}_2) \geq f(\mathbf{x}_1)$. The function f is said to be *strictly pseudoconvex* on S if whenever $\mathbf{x}_1$ and $\mathbf{x}_2$ are distinct points in S with $\nabla f(\mathbf{x}_1)^t(\mathbf{x}_2 - \mathbf{x}_1) \geq 0$, we have $f(\mathbf{x}_2) > f(\mathbf{x}_1)$.

The above generalizations of convexity extend to the concave case by replacing f by $-f$. Figure B.4 illustrates these concepts. Figure B.5 summarizes the relationships among different types of convexity.

We now give a summary of important properties for various types of convex functions. Here $f: S \to R$, where S is a nonempty convex set in R^n.

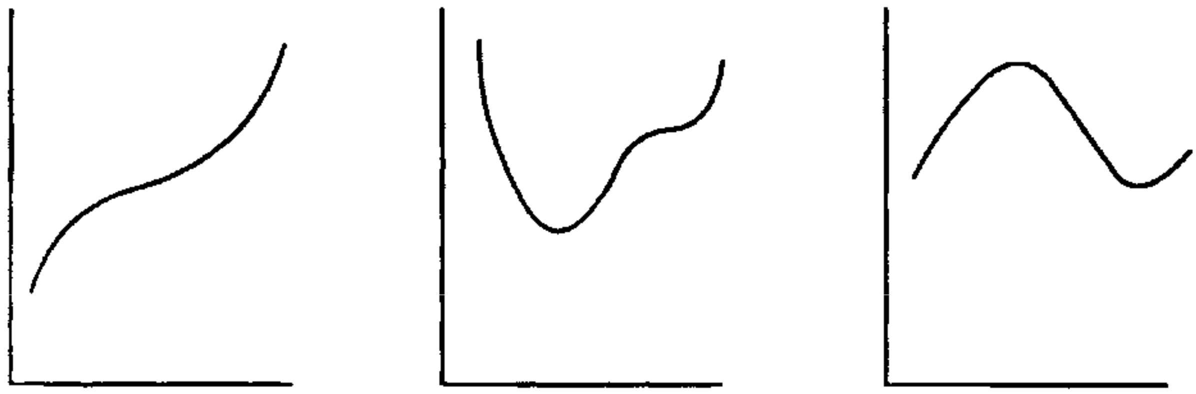


Figure B.4 Quasiconvexity and pseudoconvexity.

Strictly Convex Functions

1. The function f is continuous on the interior of S.
2. The set $\{(\mathbf{x}, y) : \mathbf{x} \in S, y \geq f(\mathbf{x})\}$ is convex.
3. The set $\{\mathbf{x} \in S : f(\mathbf{x}) \leq \alpha\}$ is convex for each real α.
4. A differentiable function f is strictly convex on S if and only if $f(\mathbf{x}) > f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}})$ for each distinct $\mathbf{x}, \bar{\mathbf{x}} \in S$.
5. Let f be twice differentiable. Then if the Hessian $\mathbf{H}(\mathbf{x})$ is positive definite for each $\mathbf{x} \in S$, f is strictly convex on S. Furthermore, if f is strictly convex on S, then the Hessian $\mathbf{H}(\mathbf{x})$ is positive semidefinite for each $\mathbf{x} \in S$.
6. Every local minimum of f over a convex set $X \subseteq S$ is the unique global minimum.
7. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, then $\bar{\mathbf{x}}$ is the unique global minimum of f over S.
8. The maximum of f over a nonempty compact polyhedral set $X \subseteq S$ is achieved at an extreme point of X.

Convex Functions

1. The function f is continuous on the interior of S.
2. The function f is convex on S if and only if the set $\{(\mathbf{x}, y) : \mathbf{x} \in S, y \geq f(\mathbf{x})\}$ is convex.
3. The set $\{\mathbf{x} \in S : f(\mathbf{x}) \leq \alpha\}$ is convex for each real α.
4. A differentiable function f is convex on S if and only if $f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}})$ for each $\mathbf{x}, \bar{\mathbf{x}} \in S$.
5. A twice differentiable function f is convex on S if and only if the Hessian $\mathbf{H}(\mathbf{x})$ is positive semidefinite for each $\mathbf{x} \in S$.
6. Every local minimum of f over a convex set $X \subseteq S$ is a global minimum.
7. If $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, then $\bar{\mathbf{x}}$ is a global minimum of f over S.

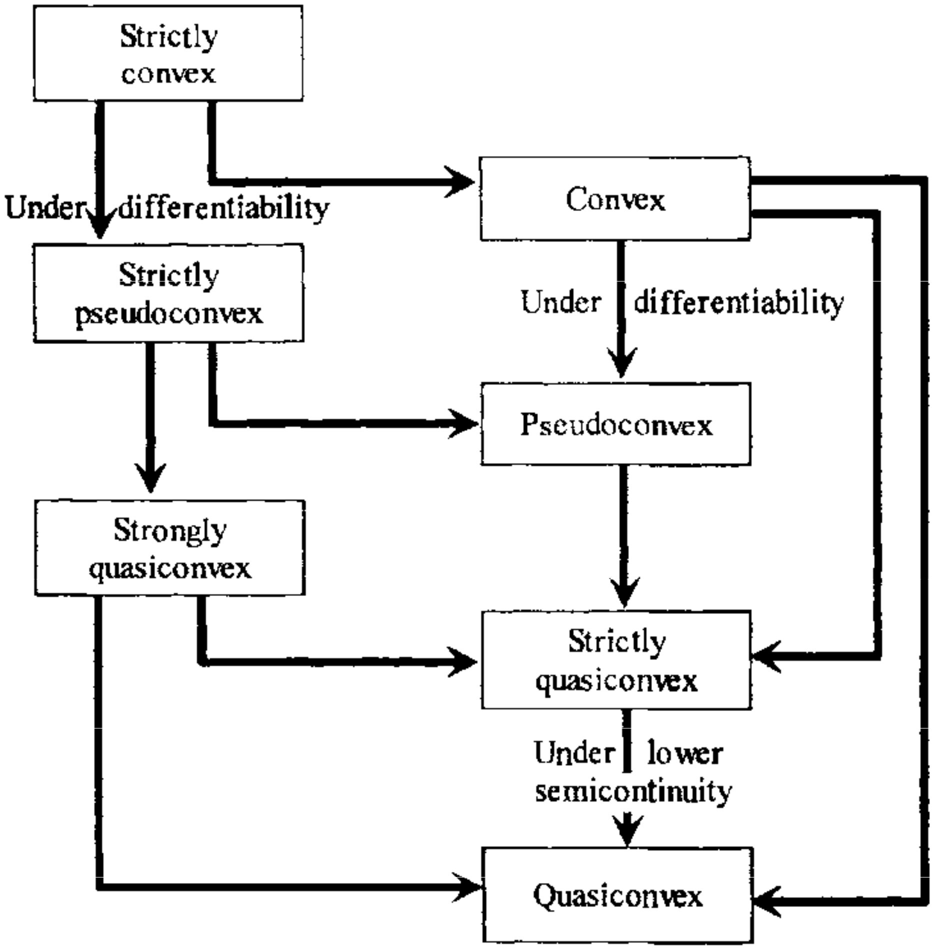


Figure B.5 Relationship among various types of convexity.

8. A maximum of f over a nonempty compact polyhedral set $X \subseteq S$ is achieved at an extreme point of X.

Pseudoconvex Functions

1. The set $\{\mathbf{x} \in S : f(\mathbf{x}) \le \alpha\}$ is convex for each real α.
2. Every local minimum of f over a convex set $X \subseteq S$ is a global minimum.
3. If $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$, then $\overline{\mathbf{x}}$ is a global minimum of f over S.
4. A maximum of f over a nonempty compact polyhedral set $X \subseteq S$ is achieved at an extreme point of X.
5. This characterization and the next relate to twice differentiable functions f defined on an open convex set $S \subseteq R^n$, with Hessian $\mathbf{H}(\mathbf{x})$. The function f is pseudoconvex on S if $\mathbf{H}(\mathbf{x}) + r(\mathbf{x})\nabla f(\mathbf{x})\nabla f(\mathbf{x})^t$ is positive semidefinite for all $\mathbf{x} \in S$, where $r(\mathbf{x}) = (1/2)[\delta - f(\mathbf{x})]$ for some $\delta > f(\mathbf{x})$. Moreover, this condition is both necessary and sufficient if f is quadratic.
6. Define the $(n + 1) \times (n + 1)$ *bordered Hessian* $\mathbf{B}(\mathbf{x})$ of f as follows, where $\mathbf{H}(\mathbf{x})$ is "bordered" by an additional row and column:

$$\mathbf{B}(\mathbf{x}) = \begin{bmatrix} \mathbf{H}(\mathbf{x}) & \nabla f(\mathbf{x}) \\ \nabla f(\mathbf{x})^t & 0 \end{bmatrix}.$$

Given any $k \in \{1,..., n\}$,and $\gamma = \{i_1,..., i_k\}$ composed of some k distinct indices $1 \le i_1 < i_2 < \cdots < i_k \le n$, the *principal submatrix* $\mathbf{B}_{\gamma,k}(\mathbf{x})$ is a $(k + 1) \times (k + 1)$ submatrix of $\mathbf{B}(\mathbf{x})$ formed by picking the elements of $\mathbf{B}(\mathbf{x})$ that intersect in the rows $i_1,..., i_k$, $(n + 1)$ and the columns $i_1,...,i_k$, $(n + 1)$ of $\mathbf{B}(\mathbf{x})$. The *leading principal submatrix* of $\mathbf{B}(\mathbf{x})$ is denoted by $\mathbf{B}_k(\mathbf{x})$ and equals $\mathbf{B}_{\gamma,k}$ for $\gamma \equiv \{1,..., k\}$. Similarly, let $\mathbf{H}_{\gamma,k}(\mathbf{x})$ and $\mathbf{H}_k(\mathbf{x})$ be the $k \times k$ principal submatrix and the leading principal submatrix, respectively, of $\mathbf{H}(\mathbf{x})$. Then f is pseudoconvex on S if for each $\mathbf{x} \in S$, we have (i) det $\mathbf{B}_{\gamma,k}(\mathbf{x}) \le 0$ for all γ, $k = 1,..., n$, and (ii) if det $\mathbf{B}_{\gamma,k}(\mathbf{x}) = 0$ for any γ, k, then det $\mathbf{H}_{\gamma,k} \ge 0$ over some neighborhood of $\mathbf{x}$. Moreover, if f is quadratic, then these conditions are both necessary and sufficient. Also, in general, the condition det $\mathbf{B}_k(\mathbf{x}) < 0$ for all $k = 1,..., n$, $\mathbf{x} \in S$, is sufficient for f to be pseudoconvex on S.

7. Let f: $S \subseteq R^n \to R$ be quadratic, where S is a convex subset of R^n. Then [f is pseudoconvex on S] $\Leftrightarrow$ [the bordered Hessian $\mathbf{B}(\mathbf{x})$ has exactly one simple negative eigenvalue for all $\mathbf{x} \in S$] $\Leftrightarrow$ [for each y $\in R^n$ such that $\nabla f(\mathbf{x})^t \mathbf{y} = 0$, we have that $\mathbf{y}^t\mathbf{H}(\mathbf{x})\mathbf{y} \ge 0$ for all $\mathbf{x} \in S$]. Moreover, [f is strictly pseudoconvex on S] $\Leftrightarrow$ [for all $\mathbf{x} \in S$, and for all $k = 1,..., n$, we have (i) det $\mathbf{B}_k(\mathbf{x}) \le 0$, and (ii) if det $\mathbf{B}_k(\mathbf{x}) = 0$, then det $\mathbf{H}_k > 0$].

Quasiconvex Functions

1. The function f is quasiconvex over S if and only if $\{\mathbf{x} \in S : f(\mathbf{x}) \le \alpha\}$ is convex for each real α.
2. A maximum of f over a nonempty compact polyhedral set $X \subseteq S$ is achieved at an extreme point of X.
3. A differentiable function f on S is quasiconvex over S if and only if $\mathbf{x}_1$, $\mathbf{x}_2 \in S$ with $f(\mathbf{x}_1) \le f(\mathbf{x}_2)$ implies that $\nabla f(\mathbf{x}_2)^t(\mathbf{x}_1 - \mathbf{x}_2) \le 0$.
4. Let f: $S \subseteq R^n \to R$, where f is twice differentiable and S is a *solid* (i.e., has a nonempty interior) convex subset of R^n. Define the bordered Hessian of f and its submatrices as in Property 6 of pseudoconvex functions. Then a sufficient condition for f to be quasiconvex on S is that for each $\mathbf{x} \in S$, det $\mathbf{B}_k(\mathbf{x}) < 0$ for all $k = 1,...,$ n. (Note that this condition actually implies that f is pseudoconvex.)

On the other hand, a necessary condition for f to be quasiconvex on S is that for each $\mathbf{x} \in S$, $\det \mathbf{B}_k(\mathbf{x}) \le 0$ for all $k = 1,\ldots, n$.

5. Let $f\colon S \subseteq R^n \to R$ be a quadratic function where $S \subseteq R^n$ is a solid (nonempty interior) convex subset of R^n. Then f is quasiconvex on S if and only if f is pseudoconvex on int(S).

A local minimum of a strictly quasiconvex function over a convex set $X \subseteq S$ is also a global minimum. Furthermore, if the function is strongly quasiconvex, the minimum is unique. If a function f is both strictly quasiconvex and lower semicontinuous, it is quasiconvex, so that the above properties for quasiconvexity hold true.

B.3 Optimality Conditions

Consider the following problem:

$$\begin{aligned} \text{P}: \quad &\text{Minimize} && f(\mathbf{x}) \\ &\text{subject to} && g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1,\ldots,m \\ &&& h_i(\mathbf{x}) = 0 \quad \text{for } i = 1,\ldots,\ell \\ &&& \mathbf{x} \in X, \end{aligned}$$

where f, g_i, h_i: $R^n \to R$ and X is a nonempty open set in R^n. We give below the *Fritz John necessary optimality conditions*. If a point $\bar{\mathbf{x}}$ is a local optimal solution to the above problem, then there must exist a nonzero vector $(u_0, \mathbf{u}, \mathbf{v})$ such that

$$u_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots, m$$

$$u_0 \ge 0,\ u_i \ge 0 \qquad \text{for } i = 1,\ldots, m,$$

where $\mathbf{u}$ and $\mathbf{v}$ are m- and ℓ-vectors whose ith components are u_i and v_i, respectively. Here, u_0, u_i, and v_i are referred to as the *Lagrange* or *Lagrangian multipliers* associated, respectively, with the objective function, the ith inequality constraint $g_i(\mathbf{x}) \le 0$, and the ith equality constraint $h_i(\mathbf{x}) = 0$. The condition $u_i g_i(\bar{\mathbf{x}}) = 0$ is called the *complementary slackness condition* and stipulates that either $u_i = 0$ or $g_i(\bar{\mathbf{x}}) = 0$. Thus, if $g_i(\bar{\mathbf{x}}) < 0$, then $u_i = 0$. By letting I be the set of binding inequality constraints at $\bar{\mathbf{x}}$, that is, $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$, then the Fritz John conditions can be written in the following equivalent form. If $\bar{\mathbf{x}}$ is a local optimal solution to Problem P above, then there exists a nonzero vector $(u_0, \mathbf{u}_I, \mathbf{v})$ satisfying the following, where $\mathbf{u}_I$ is the vector of Lagrange multipliers associated with $g_i(\mathbf{x}) \le 0$ for $i \in I$:

$$u_0 \nabla f(\overline{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\overline{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{\mathbf{x}}) = \mathbf{0}$$

$$u_0 \geq 0,\ u_i \geq 0 \qquad \text{for } i \in I.$$

If $u_0 = 0$, the Fritz John conditions become less meaningful, since essentially, they simply state that the gradients of the binding inequality constraints and the gradients of the equality constraints are linearly dependent. Under suitable assumptions, referred to as *constraint qualifications,* u_0 is guaranteed to be positive, and the Fritz John conditions reduce to the Karush–Kuhn–Tucker (KKT) conditions. A typical constraint qualification is that the gradients of the inequality constraints for $i \in I$ and the gradients of the equality constraints at $\overline{\mathbf{x}}$ are linearly independent.

The KKT necessary optimality conditions can be stated as follows. If $\overline{\mathbf{x}}$ is a local optimal solution to Problem P, under a suitable constraint qualification, there exists a vector (**u**, v) such that

$$\nabla f(\overline{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{\mathbf{x}}) = \mathbf{0}$$

$$u_i g_i(\overline{\mathbf{x}}) = 0 \qquad \text{for } i = 1,\ldots, m$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots, m.$$

Again, u_i and v_i are the *Lagrange* or *Lagrangian multipliers* associated with the constraints $g_i(\mathbf{x}) \leq 0$ and $h_i(\mathbf{x}) = 0$, respectively. Furthermore, $u_i g_i(\overline{\mathbf{x}}) = 0$ is referred to as a *complementary slackness condition.* If we let $I = \{i : g_i(\overline{\mathbf{x}}) = 0\}$, the above conditions can be rewritten as

$$\nabla f(\overline{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{\mathbf{x}}) = \mathbf{0}$$

$$u_i \geq 0 \qquad \text{for } i \in I.$$

Under suitable convexity assumptions, the KKT conditions are also *sufficient* for optimality. In particular, suppose that $\overline{\mathbf{x}}$ is a feasible solution to Problem P and that the KKT conditions stated below hold true:

$$\nabla f(\overline{\mathbf{x}}) + \sum_{i \in I} u_i \nabla g_i(\overline{\mathbf{x}}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{\mathbf{x}}) = \mathbf{0}$$

$$u_i \geq 0 \qquad \text{for } i \in I,$$

where $I = \{i : g_i(\overline{\mathbf{x}}) = 0\}$. If f is pseudoconvex, g_i is quasiconvex for $i \in I$; and if h_i is quasiconvex if $v_i > 0$ and quasiconcave if $v_i < 0$, then $\overline{\mathbf{x}}$ is an optimal solution to Problem P.

To illustrate the KKT conditions, consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & (x_1-3)^2+(x_2-2)^2 \\ \text{subject to} \quad & x_1^2+x_2^2 \le 5 \\ & x_1+2x_2 \le 4 \\ & -x_1 \le 0 \\ & -x_2 \le 0. \end{aligned}$$

The problem is illustrated in Figure B.6. Note that the optimal solution is $\bar{\mathbf{x}} = (2, 1)^t$. We first verify that the KKT conditions hold true at $\bar{\mathbf{x}}$. Here, the set of binding inequality constraints is $I = \{1, 2\}$, so that we must have $u_3 = u_4 = 0$ to satisfy the complementary slackness conditions. Note that

$$\nabla f(\bar{\mathbf{x}}) = (-2, -2)^t, \quad \nabla g_i(\bar{\mathbf{x}}) = (4, 2)^t, \quad \text{and} \quad \nabla g_2(\bar{\mathbf{x}}) = (1, 2)^t.$$

Thus, $\nabla f(\bar{\mathbf{x}}) + u_1\nabla g_1(\bar{\mathbf{x}}) + u_2\nabla g_2(\bar{\mathbf{x}}) = 0$ holds true by letting $u_1 = 1/3$ and $u_2 = 2/3$, so that the KKT conditions are satisfied at $\bar{\mathbf{x}}$. Noting that f, g_1, and g_2 are convex, we have that $\bar{\mathbf{x}}$ is indeed an optimal solution by the consequent sufficiency of the KKT conditions.

Now, let us check whether the KKT conditions hold true at the point $\hat{\mathbf{x}} = (0, 0)^t$. Here, $I = \{3, 4\}$, so that we must have $u_1 = u_2 = 0$ to satisfy complementary slackness. Note that

$$\nabla f(\hat{\mathbf{x}}) = (-6, -4)^t, \quad \nabla g_3(\hat{\mathbf{x}}) = (-1, 0)^t, \quad \text{and} \quad \nabla g_4(\hat{\mathbf{x}}) = (0, -1)^t.$$

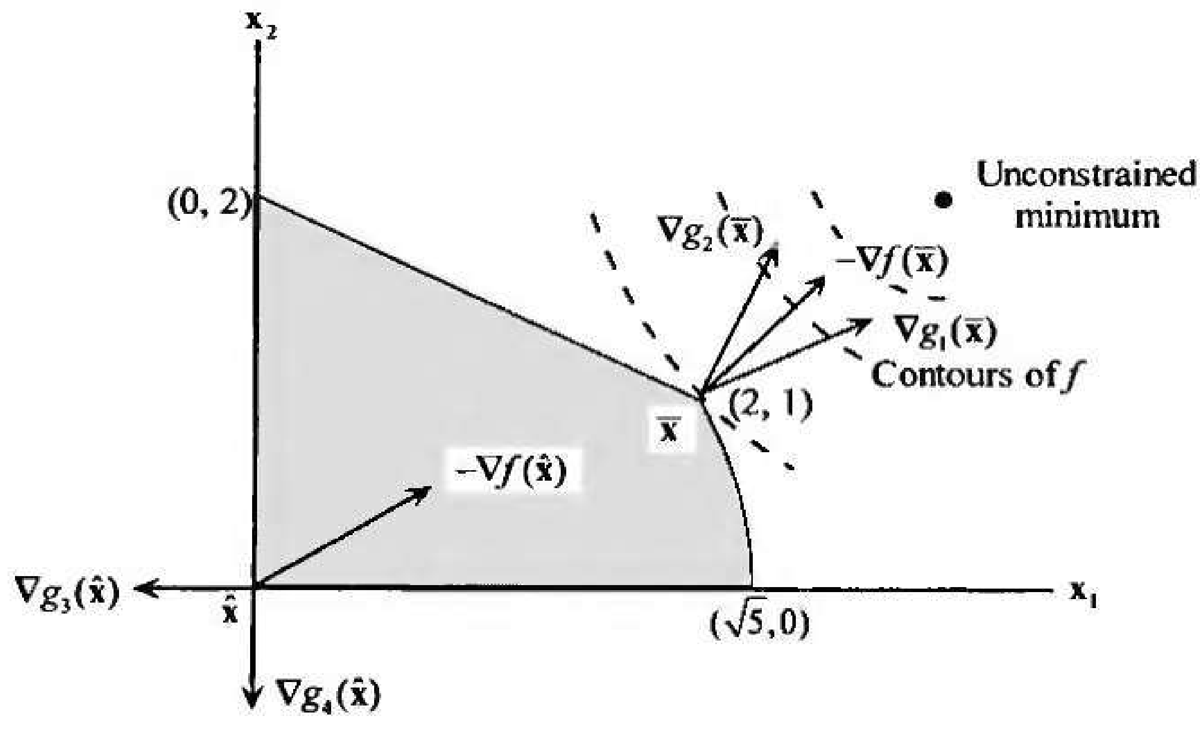


Figure B.6 The KKT conditions.

Thus, $\nabla f(\hat{\mathbf{x}}) + u_3\nabla g_3(\hat{\mathbf{x}}) + u_4\nabla g_4(\hat{\mathbf{x}}) = \mathbf{0}$ holds true only by letting $u_3 = -6$ and $u_4 = -4$, violating the nonnegativity of the Lagrange multipliers. This shows that $\hat{\mathbf{x}}$ is not a KKT point and hence could not be a candidate for an optimal solution.

In Figure B.6, the gradients of the objective function and the binding constraints are illustrated for both $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}$. Note that $-\nabla f(\bar{\mathbf{x}})$ lies in the cone spanned by the gradients of the binding constraints at $\bar{\mathbf{x}}$, whereas $-\nabla f(\hat{\mathbf{x}})$ does not lie in the corresponding cone. Indeed, the KKT conditions for a problem having inequality constraints could be interpreted geometrically as follows. A vector $\bar{\mathbf{x}}$ is a KKT point if and only if $-\nabla f(\bar{\mathbf{x}})$ lies in the cone spanned by the gradients of the binding constraints at $\bar{\mathbf{x}}$.

Let Problem P be as defined above, where all objective and constraint functions are continuously twice differentiable, and let $\bar{\mathbf{x}}$ be a KKT solution having associated Lagrange multipliers $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Define the (restricted) Lagrangian function $L(\mathbf{x}) = f(\mathbf{x}) + \bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x})$, and let $\nabla^2 L(\bar{\mathbf{x}})$ denote its Hessian at $\bar{\mathbf{x}}$. Let C denote the cone $\{\mathbf{d} : \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} = 0 \text{ for all } i \in I^+, \nabla g_i(\bar{\mathbf{x}})^t\mathbf{d} \le 0 \text{ for all } i \in I^0, \text{ and } \nabla h_i(\bar{\mathbf{x}})^t\mathbf{d} = 0 \text{ for all } i = 1,\ldots,\ell\}$, where $I^+ = \{i \in \{1,\ldots, m\} : \bar{u}_i > 0\}$ and $I^0 = \{1,\ldots, m\} - I^+$. Then we have the following *second-order sufficient conditions* holding true: If $\nabla^2 L(\bar{\mathbf{x}})$ is positive definite on C, that is, $\mathbf{d}^t\nabla^2 L(\bar{\mathbf{x}})\mathbf{d} > 0$ for all $\mathbf{d} \in C$, $\mathbf{d} \ne \mathbf{0}$, then $\bar{\mathbf{x}}$ is a strict local minimum for Problem P. We also remark that if $\nabla^2 L(\mathbf{x})$ is positive semidefinite for all feasible $\mathbf{x}$ [respectively, for all feasible $\mathbf{x}$ in $N_\varepsilon(\bar{\mathbf{x}})$ for some $\varepsilon > 0$], then $\bar{\mathbf{x}}$ is a global (respectively, local) minimum for P.

Conversely, suppose that $\bar{\mathbf{x}}$ is a local minimum for P, and let the gradients $\nabla g_i(\bar{\mathbf{x}})$, $i \in I$, $\nabla h_i(\bar{\mathbf{x}})$, $i = 1,\ldots, \ell$ be linearly independent, where $I = \{i \in \{1,\ldots, m\} : g_i(\bar{\mathbf{x}}) = 0\}$. Define the cone C as stated above for the second-order sufficiency conditions. Then $\bar{\mathbf{x}}$ is a KKT point having associated Lagrange multipliers $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Moreover, defining the (restricted) Lagrangian function $L(\mathbf{x}) = f(\mathbf{x}) + \bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x})$, the *second-order necessary condition* is that $\nabla^2 L(\bar{\mathbf{x}})$ is positive semidefinite on C.

B.4 Lagrangian Duality

Given a nonlinear programming problem, called the *primal problem*, there exists a problem that is closely associated with it, called the *Lagrangian dual problem*. These two problems are given below.

$$\begin{aligned} \text{Primal Problem P: Minimize } & f(\mathbf{x}) \\ \text{subject to } & g_i(\mathbf{x}) \le 0 \quad \text{for } i = 1,\dots,m \\ & h_i(\mathbf{x}) = 0 \quad \text{for } i = 1,\dots,\ell \\ & \mathbf{x} \in X, \end{aligned}$$

where f, g_i, and h_i: $R^n \to R$ and X is a nonempty set in R^n. Let **g** and **h** be the m- and ℓ-vector functions whose ith components are, respectively, g_i and h_i.

$$\begin{aligned} \text{Lagrangian Dual Problem D: Maximize } & \theta(\mathbf{u}, \mathrm{v}) \\ \text{subject to } & \mathbf{u} \ge \mathbf{0}, \end{aligned}$$

where $\theta(\mathbf{u}, \mathrm{v}) = \inf\{f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) + \sum_{i=1}^{\ell} v_i h_i(\mathbf{x}) : \mathbf{x} \in X\}$. Here the vectors **u** and v belong to R^m and R^ℓ, respectively. The ith component u_i of **u** is referred to as the dual variable or Lagrange/Lagrangian multiplier associated with the constraint $g_i(\mathbf{x}) \le 0$, and the ith component v_i of v is referred to as the dual variable or Lagrange/Lagrangian multiplier associated with the constraint $h_i(\mathbf{x}) = 0$. It may be noted that θ is a *concave* function, even in the absence of any convexity or concavity assumptions on f, g_i, or h_i, or convexity of the set X.

We summarize below some important relationships between the primal and dual problems:

1. If **x** is feasible to Problem P and if (**u**, v) is feasible to Problem D, then $f(\mathbf{x}) \ge \theta(\mathbf{u}, \mathrm{v})$. Thus,

   $$\inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{x} \in X\} \ge \sup\{\theta(\mathbf{u}, \mathrm{v}) : \mathbf{u} \ge \mathbf{0}\}.$$

   This result is called the *weak duality theorem.*
2. If $\sup\{\theta(\mathbf{u}, \mathrm{v}) : \mathbf{u} \ge \mathbf{0}\} = \infty$, then there exists no point $\mathbf{x} \in X$ such that $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, so that the primal problem is infeasible.
3. If $\inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{x} \in X\} = -\infty$, then $\theta(\mathbf{u}, \mathrm{v}) = -\infty$ for each (**u**, v) with $\mathbf{u} \ge \mathbf{0}$.
4. If there exists a feasible **x** to the primal problem and a feasible (**u**, v) to the dual problem such that $f(\mathbf{x}) = \theta(\mathbf{u}, \mathrm{v})$, then **x** is an optimal solution to Problem P and (**u**, v) is an optimal solution to Problem D. Furthermore, the complementary slackness condition $u_i g_i(\mathbf{x}) = 0$ for $i = 1,\dots, m$ holds true.
5. Suppose that X is convex, that f, g_i: $R^n \to R$ for $i = 1,\dots, m$ are convex, and that **h** is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$, where **A** is an $m \times n$ matrix and **b** is an m-vector. Under a suitable constraint qualification, the optimal objective values of Problems P and D are equal; that is,

   $$\inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\} = \sup\{\theta(\mathbf{u}, \mathrm{v}) : \mathbf{u} \ge \mathbf{0}\}.$$

Furthermore, if the inf is finite, then the sup is achieved at $(\overline{\mathbf{u}}, \overline{\mathbf{v}})$ with $\overline{\mathbf{u}} \geq \mathbf{0}$. Also, if the inf is achieved at $\overline{\mathbf{x}}$, then $u_i g_i(\overline{\mathbf{x}}) = 0$ for $i = 1,\ldots, m$. This result is referred to as the *strong duality theorem*.