Nonlinear Programming: Theory and Algorithms
by Mokhtar S. Bazaraa, Hanif D. Sherali and C. M. Shetty

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):

$$\text{BLP}: \text{ Minimize } \left\{ \sum_{j=1}^{p} y_j w_j + (1-y_j) z_j : \right. \\ \left. \mathbf{w} - \mathbf{Mz} = \mathbf{q}, \mathbf{w} \geq \mathbf{0}, \mathbf{z} \geq \mathbf{0}, \quad \text{and } \mathbf{y} \text{ binary} \right\}. \tag{11.4}$$

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

$$h(\mathbf{y}) \equiv \min\{\sum_{j=1}^{n} [y_j w_j + (1-y_j)z_j] : \mathbf{w} - \mathbf{Mz} = \mathbf{q},\ \mathbf{w} \ge \mathbf{0}, \text{ and } \mathbf{z} \ge \mathbf{0}\}$$

(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 $\mathbf{q} \ngeq \mathbf{0}$, however, a new column **e** and an artificial variable are introduced, leading to the following system, where **e** is a vector of p ones.

$$\mathbf{w} - \mathbf{Mz} - \mathbf{e}z_0 = \mathbf{q} \tag{11.5}$$

$$z_0 \ge 0,\ w_j \ge 0,\ z_j \ge 0 \qquad \text{for } j = 1,\dots,p \tag{11.6}$$

$$w_j z_j = 0 \qquad \text{for } j = 1,\dots,p. \tag{11.7}$$

Letting $z_0 = \text{maximum } \{-q_i : 1 \le i \le p\}$, $\mathbf{z} = \mathbf{0}$, 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 $(\mathbf{w}, \mathbf{z}, z_0)$ to this system is called an *almost complementary basic feasible solution* if:

1. $(\mathbf{w}, \mathbf{z}, z_0)$ 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 $(\mathbf{w}, \mathbf{z}, z_0)$, where w_s and z_s are both nonbasic, an *adjacent almost complementary basic feasible solution* $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0)$ 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 $\mathbf{M}$, the algorithm converges in a finite number of steps, with a complementary basic feasible solution.

Initialization Step If $\mathbf{q} \geq \mathbf{0}$, stop; $(\mathbf{w}, \mathbf{z}) = (\mathbf{q}, \mathbf{0})$ 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 $\mathbf{d}_s \leq \mathbf{0}$, go to Step 4. Otherwise, determine the index r by

the following minimum ratio test, where $\bar{\mathbf{q}}$ is the updated right-hand side column denoting the values of the basic variables:

$$\frac{\bar{q}_r}{d_{rs}} = \min_{1 \le i \le p} \left\{ \frac{\bar{q}_i}{d_{is}} : d_{is} > 0 \right\}.$$

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 $R = \{(\mathbf{w}, \mathbf{z}, z_0) + \lambda \mathbf{d} : \lambda \ge 0\}$ 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 $j = 1,\ldots,4$.

11.1.4 Example (Ray Termination)

We wish to find a solution to the linear complementary problem defined by

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 2 \end{bmatrix}, \qquad \mathbf{q} = \begin{bmatrix} 1 \\ 4 \\ -2 \\ -4 \end{bmatrix}.$$

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 $\mathbf{M}$, 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. $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0) \neq (\mathbf{0}, \mathbf{0}, 0)$, $(\hat{\mathbf{w}}, \hat{\mathbf{z}}) \geq \mathbf{0}$, $\hat{z}_0 \geq 0$.

---

* 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 $(\hat{\mathbf{w}}, \hat{\mathbf{z}}, \hat{z}_0) \neq (\mathbf{0}, \mathbf{0}, 0)$. 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 $\mathbf{z} \geq \mathbf{0}$ 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 $\mathbf{M}$ 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 $\mathbf{M}$ 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 $\mathbf{M}$ 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 $\mathbf{u}$ and $\mathbf{v}$, respectively, and denoting the vector of slack variables by $\mathbf{y}$, 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 $\mathbf{y}$ and the Lagrangian multiplier vectors for the constraints $\mathbf{Ax} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ by $\mathbf{u}$ and $\mathbf{v}$, 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 $\mathbf{H}$ 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

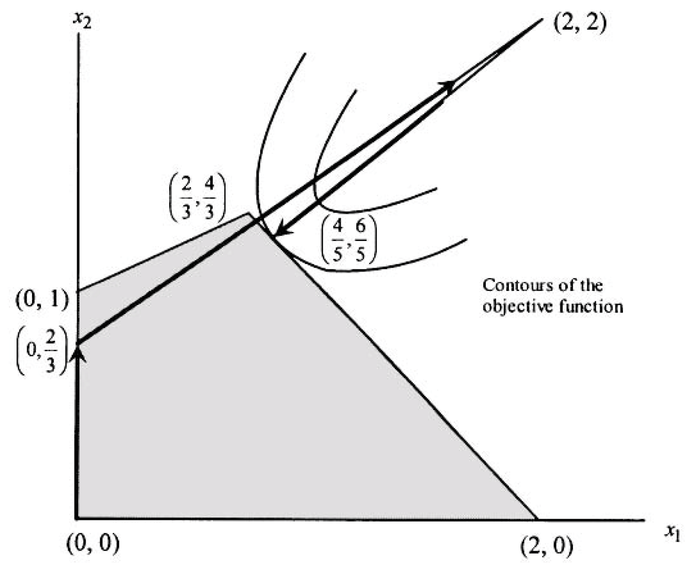


Figure 11.1 Points generated by the complementary pivoting algorithm.

$$\mathbf{A}=\begin{bmatrix}-1 & 1\\ 1 & -2\end{bmatrix},\quad \mathbf{H}=\begin{bmatrix}2 & -2\\ -2 & 2\end{bmatrix},\quad \mathbf{b}=\begin{bmatrix}1\\ 4\end{bmatrix},\quad \text{and}\quad \mathbf{c}=\begin{bmatrix}-2\\ -4\end{bmatrix}.$$

Denote the vector of slacks by **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

$$\mathbf{M}=\begin{bmatrix}0 & 0 & 1 & -1\\ 0 & 0 & -1 & 2\\ -1 & 1 & 2 & -2\\ 1 & -2 & -2 & 2\end{bmatrix}\quad\text{and}\quad \mathbf{q}=\begin{bmatrix}1\\ 4\\ -2\\ -4\end{bmatrix}.$$

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 **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 **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 **A** be an $m\times n$ matrix, and let **H** be an $n\times n$ symmetric matrix. If $\mathbf{y}^t\mathbf{Hy}\ge 0$ for each $\mathbf{y}\ge\mathbf{0}$, then the matrix

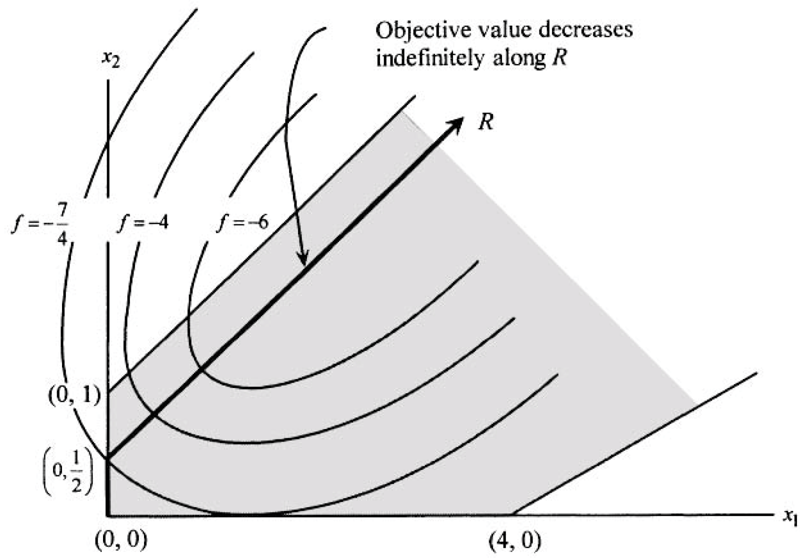


Figure 11.2 Unbounded optimal solution and ray termination.

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & -\mathbf{A} \\ \mathbf{A}^t & \mathbf{H} \end{bmatrix}$$

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

$$(\mathbf{M}+\mathbf{M}^t)\mathbf{z}=\begin{bmatrix}\mathbf{0}\\ 2\mathbf{Hy}\end{bmatrix}.$$

Since $\mathbf{z}^t\mathbf{Mz}=0$, we get $\mathbf{y}^t\mathbf{Hy}=0$ by (11.12). By assumption, since $\mathbf{y}\geq\mathbf{0}$ and $\mathbf{y}^t\mathbf{Hy}=0$, we get $\mathbf{Hy}=\mathbf{0}$, 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 $\mathbf{y}^t\mathbf{Hy}=0$ implies that $\mathbf{Hy}=\mathbf{0}$, so $\mathbf{M}$ is copositive plus.

Proof

It suffices to show that $\mathbf{y}^t\mathbf{Hy}=0$ implies that $\mathbf{Hy}=\mathbf{0}$. Let $\mathbf{Hy}=\mathbf{d}$, and noting that $\mathbf{H}$ is positive semidefinite, we get

$$0\leq(\mathbf{y}^t-\lambda\mathbf{d}^t)\mathbf{H}(\mathbf{y}-\lambda\mathbf{d})=\mathbf{y}^t\mathbf{Hy}+\lambda^2\mathbf{d}^t\mathbf{Hd}-2\lambda\|\mathbf{d}\|^2.$$

Since $\mathbf{y}^t\mathbf{Hy}=0$, dividing the above inequality by λ and letting $\lambda\to 0^+$, it follows that $\mathbf{0}=\mathbf{d}=\mathbf{Hy}$.

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 $\mathbf{y}^t\mathbf{Hy}=0$, then $\mathbf{y}=\mathbf{0}$, and hence $\mathbf{Hy}=\mathbf{0}$. By the theorem, $\mathbf{M}$ is copositive plus.

11.2.4 Theorem

Consider the problem to minimize $\mathbf{c}^t\mathbf{x}+(1/2)\mathbf{x}^t\mathbf{Hx}$ subject to $\mathbf{Ax}\leq\mathbf{b}$, $\mathbf{x}\geq\mathbf{0}$. 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 $\mathbf{w}-\mathbf{Mz}=\mathbf{q}$, $(\mathbf{w},\mathbf{z})\geq\mathbf{0}$, $\mathbf{w}^t\mathbf{z}=0$, 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

$$0 > \mathbf{c}^t\mathbf{d} + \mathbf{b}^t\mathbf{f} = \mathbf{c}^t\mathbf{d} + (\hat{\mathbf{y}} + \mathbf{A}\hat{\mathbf{x}})^t\mathbf{f} \ge \mathbf{c}^t\mathbf{d} + \hat{\mathbf{x}}^t\mathbf{A}^t\mathbf{f} \ge \mathbf{c}^t\mathbf{d} + \hat{\mathbf{x}}^t\mathbf{Hd}. \tag{11.14b}$$

Now, suppose that $\mathbf{H}$ is positive semidefinite. By (11.14a) it follows that $\mathbf{d}^t\mathbf{Hd} = 0$, and by Corollary 1 to Theorem 11.2.3 it follows that $\mathbf{Hd} = \mathbf{0}$. By (11.14b) we have $\mathbf{c}^t\mathbf{d} < 0$. Since $\mathbf{Ad} \le \mathbf{0}$ and $\mathbf{d} \ge \mathbf{0}$, $\mathbf{d}$ 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 $f(\mathbf{x}) = \mathbf{c}^t\mathbf{x} + (1/2)\mathbf{x}^t\mathbf{Hx}$. Since $\mathbf{Hd} = \mathbf{0}$, we get

$$f(\hat{\mathbf{x}} + \lambda\mathbf{d}) = f(\hat{\mathbf{x}}) + \lambda(\mathbf{c}^t + \hat{\mathbf{x}}^t\mathbf{H})\mathbf{d} + \frac{1}{2}\lambda^2\mathbf{d}^t\mathbf{Hd} = f(\hat{\mathbf{x}}) + \lambda\mathbf{c}^t\mathbf{d}.$$

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), $\mathbf{d}^t\mathbf{Hd} \le 0$. 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 $\mathbf{Hd} = \mathbf{0}$ 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 $\mathbf{H}$ 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:

$$\begin{aligned} \text{NQP:}\ \ \text{Minimize}\ \ & \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{Hx} \\ \text{subject to}\ \ & \mathbf{Ax} \le \mathbf{b} \end{aligned}$$

$$\mathbf{x} \in \Omega \equiv \{\mathbf{x} : \ell_j \le x_j \le u_j \text{ for } j = 1,\ldots, n\},$$

where $\mathbf{H}$ is $n \times n$, symmetric, but not necessarily positive semidefinite, $\mathbf{A}$ is an $m \times n$ matrix, and where the *hyperrectangle* Ω defines finite lower and upper bounds on the variables, with $\ell_j < u_j$, $\forall j = 1,\ldots, n$. 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 $\mathbf{H}$ and denoting the collection of $m + 2n$ inequalities $\mathbf{Ax} \le \mathbf{b}$, and $\ell_j \le x_j \le u_j$ for $j = 1,\ldots, n$ jointly by $\mathbf{G}_i\mathbf{x} \equiv \sum_{k=1}^n G_{ik}x_k \le g_i$, for $i = 1,\ldots, \bar{m} \equiv m + 2n$, 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 $w_{12} = x_1 x_2$ 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 $\mathbf{H}$, 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.)

$$\text{LP}(\Omega)\text{: Minimize } \sum_{k=1}^{n} c_k x_k + \frac{1}{2}\sum_{k=1}^{n} h_{kk} w_{kk} + \sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n} h_{k\ell} w_{k\ell} \qquad (11.17\text{a})$$

$$\text{subject to } [(g_i - \mathbf{G}_i\mathbf{x})(g_j - \mathbf{G}_j\mathbf{x})]_L \ge 0 \text{ for } 1 \le i \le j \le m. \qquad (11.17\text{b})$$

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 v[P].

11.2.5 Lemma

a. Let $\overline{\mathbf{x}}$ be any feasible solution to Problem NQP, and let $\overline{\mathbf{w}}$ be defined according to (11.16) (i.e., $\overline{w}_{k\ell} \equiv \overline{x}_k\overline{x}_\ell$ for $1 \le k \le \ell \le n$). Then $(\overline{\mathbf{x}}, \overline{\mathbf{w}})$ is feasible to LP(Ω) and yields the same objective value as in NQP. Hence, in particular, v[LP(Ω)] $\le v$[NQP].
b. Conversely, let $(\overline{\mathbf{x}}, \overline{\mathbf{w}})$ be any feasible solution to LP(Ω). Then $\overline{\mathbf{x}}$ 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 $(\overline{\mathbf{x}}, \overline{\mathbf{w}})$ 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 $(\overline{\mathbf{x}}, \overline{\mathbf{w}})$ is feasible to LP(Ω), we have that $\overline{\mathbf{x}}$ satisfies $g_i - \mathbf{G}_i(\mathbf{x}) \ge 0$. Therefore, $\overline{\mathbf{x}}$ 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 v[LP(Ω)]. But v[LP(Ω)] $\le v$[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 $(\overline{\mathbf{x}}, \overline{\mathbf{w}})$ be any feasible solution to LP(Ω). Suppose that $\overline{x}_p = \ell_p$ or $\overline{x}_p = u_p$ for some $p \in \{1,..., n\}$. Then $\overline{w}_{(pq)} = \overline{x}_p\overline{x}_q$ 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 $(\overline{\mathbf{x}}, \overline{\mathbf{w}}) \equiv (\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$. Compute a *discrepancy index*

$$\theta_k \equiv \max\{0,\ h_{kk}(\overline{x}_k^2 - \overline{w}_{kk})\} + \sum_{\ell=1}^{n} \max\{0,\ h_{k\ell}(\overline{x}_k \overline{x}_\ell - \overline{w}_{k\ell})\} \qquad \text{for } k = 1,\ldots,n, \tag{11.18a}$$

and determine the *branching variable* x_p, where the index p corresponds to

$$\theta_p = \max\{\theta_k, k = 1,\ldots,n\}. \tag{11.18b}$$

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 $\overline{x}_p \equiv x_p^{q(s)}$ 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:

$$\ell_p^{q(s)} \le x_p \le x_p^{q(s)} \equiv \overline{x}_p \quad \text{and} \quad \overline{x}_p \equiv x_p^{q(s)} \le x_p \le u_p^{q(s)}. \tag{11.18c}$$

Note that because $\mathrm{LB}_{q(s)} < v^*$, 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 $k = 1,\ldots, n$, so, by (11.18a), this would mean that for each $k = 1,\ldots, n$,

$$h_{kk}\overline{x}_k^2 \le h_{kk}\overline{w}_{kk} \qquad \text{and} \qquad h_{k\ell}\overline{x}_k\overline{x}_\ell \le h_{k\ell}\overline{w}_{(k\ell)} \qquad \text{for } \ell = 1,\ldots, n, \tag{11.18e}$$

or that the objective value in NQP for the solution $\overline{\mathbf{x}}$ is less than or equal to $v[\mathrm{LP}(\Omega^{q(s)})] = \mathrm{LB}_{q(s)}$, which contradicts $v^* > \mathrm{LB}_{q(s)}$. 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 LP(Ω^1) and let $(\bar{\mathbf{x}}, \bar{\mathbf{w}})$ be the solution obtained of objective value $\mathrm{LB}_1 = v[\mathrm{LP}(\Omega^1)]$. 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 $\mathrm{LB}_1 + \varepsilon \geq v^*$, 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. Bounding 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 $Q_{s+1} = Q_s - \{q \in Q_s : \mathrm{LB}_q + \varepsilon \geq v^*\}$. 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 $q(s) \in \mathrm{argmin}\{\mathrm{LB}_q : q \in Q_s\}$, 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 $\Omega^{q(s)} \equiv \{\mathbf{x} : \boldsymbol{\ell}^{q(s)} \le \mathbf{x} \le \mathbf{u}^{q(s)}\}$, denote $(\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$ as the optimal solution obtained for $\text{LP}(\Omega^{q(s)})$, and let $\theta^{q(s)} \equiv (\theta_k^{q(s)},\ k = 1,\ldots, n)$ denote the discrepancy index vector determined via (11.18a, b) for the solution $(\overline{\mathbf{x}}, \overline{\mathbf{w}}) \equiv (\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)})$. By taking any convergent subsequence, if necessary, using the boundedness of the sequences generated, assume without loss of generality that

$$\{(\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)}, \boldsymbol{\ell}^{q(s)}, \mathbf{u}^{q(s)}, \theta^{q(s)})\}_S \to (\hat{\mathbf{x}}, \hat{\mathbf{w}}, \hat{\boldsymbol{\ell}}, \hat{\mathbf{u}}, \hat{\theta}).$$

Note that by the continuity of the constraint functions in (11.17), $(\hat{\mathbf{x}}, \hat{\mathbf{w}})$ is feasible to $\text{LP}(\hat{\Omega})$, where $\hat{\Omega} \equiv \{\mathbf{x}\colon \hat{\boldsymbol{\ell}} \le \mathbf{x} \le \hat{\mathbf{u}}\}$. 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 $\{x_p^{q(s)}\} \to \overline{x}_p \in [\hat{\ell}_p, \hat{u}_p]$ because $(\hat{\mathbf{x}}, \hat{\mathbf{w}})$ is feasible to $\text{LP}(\hat{\Omega})$ 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 $\text{LP}(\hat{\Omega})$, this in turn means that $\hat{w}_{(pq)} = \hat{x}_p\hat{x}_q$, for all $q = 1,\ldots, n$. 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 $k = 1,\ldots, n$, so $0 = \hat{\theta}_p \ge \hat{\theta}_k \ge 0$ for all $k = 1,\ldots, n$ (i.e., $\hat{\theta}_k = 0$ for all $k = 1,\ldots, n$). As in (11.18e), denoting the objective function of $\text{LP}(\cdot)$ in (11.17a) by $f_{\text{LP}}(\mathbf{x}, \mathbf{w})$, this implies that

$$\mathbf{c}^t\hat{\mathbf{x}}+\frac{1}{2}\hat{\mathbf{x}}^t\mathbf{H}\hat{\mathbf{x}} \le f_{\mathrm{LP}}(\hat{\mathbf{x}}, \hat{\mathbf{w}}). \tag{11.19a}$$

But by Lemma 11.2.5 and our least lower bound node selection rule, since $f_{\mathrm{LP}}(\mathbf{x}^{q(s)}, \mathbf{w}^{q(s)}) = v[\mathrm{LP}(\Omega^{q(s)})] = \mathrm{LB}(s) \le v[\mathrm{NQP}]$, $\forall\, s \in S$, taking limits as $s \to \infty$, $s \in S$, we get

$$f_{\mathrm{LP}}(\hat{\mathbf{x}}, \hat{\mathbf{w}}) \le v[\mathrm{NQP}]. \tag{11.19b}$$

Putting (11.19a) and (11.19b) together, and noting from above that $\hat{\mathbf{x}}$ is feasible to NQP, we deduce that

$$v[\mathrm{NQP}] \le \mathbf{c}^t\hat{\mathbf{x}}+\frac{1}{2}\mathbf{x}^t\mathbf{H}\hat{\mathbf{x}} \le f_{\mathrm{LP}}(\hat{\mathbf{x}}, \hat{\mathbf{w}}) \le v[\mathrm{NQP}], \tag{11.19c}$$

so equality holds true throughout (11.19c). This means that $\hat{\mathbf{x}}$ solves NQP, and this completes the proof.

11.2.8 Example

Consider the following concave minimization quadratic programming problem:

$$\begin{aligned} \text{NQP:}\ \ \text{Minimize}\ \ & -(x_1-12)^2-x_2^2 \\ \text{subject to}\ \ & -3x_1+4x_2 \le 24 \\ & 3x_1+8x_2 \le 120 \\ \mathbf{x} \in \Omega = \{\mathbf{x} : 0 \le x_1 \le 24,\ & 0 \le x_2 \le 15\}. \end{aligned}$$

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 $\mathbf{x}$-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:

$$\begin{aligned} \text{P: Minimize } & \sum_{j=1}^{n} f_j(x_j) \\ \text{subject to } & \sum_{j=1}^{n} g_{ij}(x_j) \le p_i \quad \text{for } i=1,\ldots,m \\ & x_j \ge 0 \quad \text{for } j=1,\ldots,n. \end{aligned} \tag{11.20}$$

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

$$\hat{\theta}(\mu) = \lambda\theta(\mu_v) + (1-\lambda)\theta(\mu_{v+1}). \tag{11.21}$$

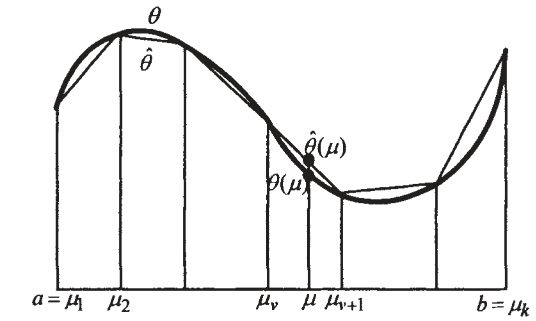


Figure 11.3 Piecewise linear approximation of a function.

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

$$\hat{\theta}(\mu) = \sum_{v=1}^{k} \lambda_v \theta(\mu_v), \quad \sum_{v=1}^{k} \lambda_v = 1, \quad \lambda_v \geq 0 \qquad \text{for } v = 1, \ldots, k \tag{11.22}$$

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.

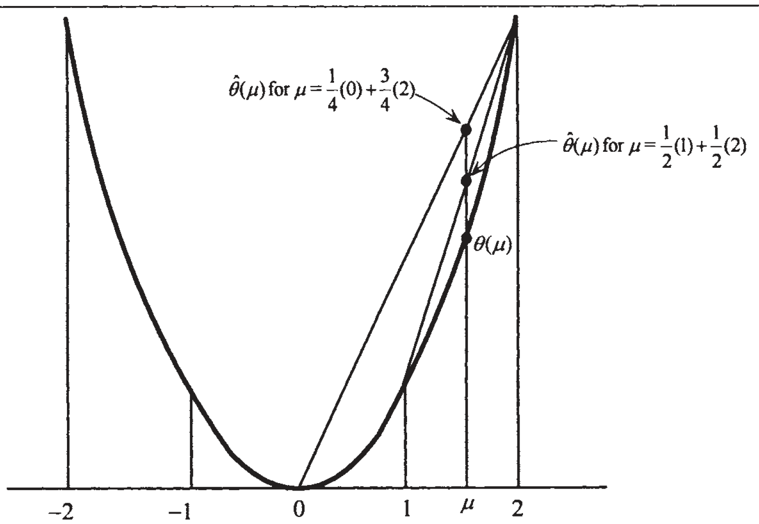


Figure 11.4 Importance of adjacency in approximation.

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 $i = 1,\dots, m$ and replacing it with the piecewise linear approximation defined by (11.22). For the sake of clarity, we define a set L as

$$L = \{j : f_j \text{ and } g_{ij} \text{ for } i = 1,\dots, m \text{ are linear}\}.$$

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 $v = 1,\dots, k_j$, 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 $i = 1,\dots, m$ could be replaced by their linear approximations

$$\hat{f}_j(x_j) = \sum_{v=1}^{k_j} \lambda_{vj} f(x_{vj}) \qquad \text{for } j \notin L$$

$$\hat{g}_{ij}(x_j) = \sum_{v=1}^{k_j} \lambda_{vj} g_{ij}(x_{vj}) \qquad \text{for } i = 1,\dots, m \text{ and } j \notin L$$

$$\sum_{v=1}^{k_j} \lambda_{vj} = 1 \qquad \text{for } j \notin L$$

$$\lambda_{vj} \geq 0 \qquad \text{for } v = 1,\dots, k_j \text{ and } j \notin L.$$

By definition, both f_j and g_{ij} for $i = 1,\dots, m$ are linear for $j \in L$. Hence, no grid points need be defined, and in this case, the linear approximations are given by

$$\hat{f}_j(x_j) \equiv f_j(x_j),\ \hat{g}_{ij}(x_j) \equiv g_{ij}(x_j) \quad \text{for } i = 1,\dots,m \text{ and } j \in L.$$

The following Problem AP can then be viewed as the problem that approximates the original Problem P.

$$\begin{aligned}
\text{AP: Minimize} \quad & \sum_{j\in L} f_j(x_j) + \sum_{j\notin L} \hat{f}_j(x_j) \\
\text{subject to} \quad & \sum_{j\in L} g_{ij}(x_j) + \sum_{j\notin L} \hat{g}_{ij}(x_j) \leq p_i \quad \text{for } i = 1,\dots,m \qquad (11.23)\\
& x_j \geq 0 \quad \text{for } j = 1,\dots,n.
\end{aligned}$$

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,..., 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,..., m, and $x_j \ge 0$ for j = 1,..., n. Let $L = \{j : f_j$ and g_{ij} for i = 1,..., m are linear$\}$. Assume that for $j \notin L$, f_j is strictly convex and that g_{ij} is convex for i = 1,..., m. Suppose further that for each $j \notin L$, f_j and g_{ij} for i = 1,..., m are replaced by their piecewise linear approximations via the grid points x_{vj} for v = 1,..., 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 $v = \gamma$. By the strict convexity of f_j, the convexity of g_{ij}, and by (11.26) and (11.27), we have

$$\begin{aligned} f_j(x_{\gamma j}) + \sum_{i=1}^{m} u_i g_{ij}(x_{\gamma i}) + v_j &< \alpha_1 f_j(x_{\ell j}) + \alpha_2 f_j(x_{pj}) \\ &+ \sum_{i=1}^{m} u_i[\alpha_1 g_{ij}(x_{\ell j}) + \alpha_2 g_{ij}(x_{pj})] + v_j = 0. \end{aligned}$$

This contradicts (11.28) for $v = \gamma$, 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 $i = 1,\dots, m$, and noting that $\hat{x}_j$ for $j \in L$, and $\hat{\lambda}_{vj}$ for $v = 1,\dots, k_j$, $j \notin L$, satisfy the constraints in (11.25), we get

$$\begin{aligned} g_i(\hat{\mathbf{x}}) &= \sum_{j\in L} g_{ij}(\hat{x}_j) + \sum_{j\notin L} g_{ij}(\hat{x}_j) \\ &= \sum_{j\in L} g_{ij}(\hat{x}_j) + \sum_{j\notin L} g_{ij}\left(\sum_{v=1}^{k_j} \hat{\lambda}_{vj} x_{vj}\right) \\ &\le \sum_{j\in L} g_{ij}(\hat{x}_j) + \sum_{j\notin L} \sum_{v=1}^{k_j} \hat{\lambda}_{vj} g_{ij}(x_{vj}) \\ &\le p_i \end{aligned}$$

for $i = 1,\dots, m$. Furthermore, $\hat{x}_j \ge 0$ for $j \in L$, and $\hat{x}_j = \sum_{v=1}^{k_j} \hat{\lambda}_{vj} x_{vj} \ge 0$ for $j \notin L$, since $\hat{\lambda}_{vj}$, $x_{vj} \ge 0$ for $v = 1,\dots, k_j$ 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:

$$\begin{aligned} \text{Minimize } & x_1^2 - 6x_1 + x_2^2 - 8x_2 - \frac{1}{2}x_3 \\ \text{subject to } & x_1 + x_2 + x_3 \le 5 \\ & x_1^2 - x_2 \le 3 \\ & x_1, x_2, x_3 \ge 0. \end{aligned}$$

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

$$\hat{g}_i(\mathbf{x}) \geq g_i(\mathbf{x}) \geq \hat{g}_i(\mathbf{x}) - \bar{c}_i \qquad \text{for } i = 1,\ldots, m. \tag{11.34}$$

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 $L = \{j : f_j \text{ and } g_{ij} \text{ for } i = 1,\ldots, m \text{ are linear}\}$. 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 $i = 1,\ldots, m$. 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 $i = 1,\ldots, m$ 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 $v = 1,\ldots, k_j, j \notin L$, 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 $i = 1,\ldots, m$. 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 $i = 1,\ldots, m$, 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 $i = 1,\ldots, m$, 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 $i = 1,\ldots, m$ 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 $i = 1,\ldots, m$. Letting $\mathbf{x} = \bar{\mathbf{x}}$ in (11.35) and noting that $\hat{\mathbf{u}} \geq \mathbf{0}$, it follows that

$$\hat{f}(\hat{\mathbf{x}}) \le \hat{f}(\overline{\mathbf{x}}) + c\sum_{i=1}^{m} \hat{u}_i. \tag{11.36}$$

By Part 1 of the theorem, $\hat{\mathbf{x}}$ is feasible to Problem P, and hence, $f(\hat{\mathbf{x}}) \ge f(\overline{\mathbf{x}})$. From Theorem 11.3.3, $f(\overline{\mathbf{x}}) \ge \hat{f}(\overline{\mathbf{x}}) - c$, and hence, $f(\hat{\mathbf{x}}) \ge f(\overline{\mathbf{x}}) \ge \hat{f}(\overline{\mathbf{x}}) - c$. From (11.36) and since $\hat{f}(\hat{\mathbf{x}}) \ge f(\hat{\mathbf{x}})$, it follows that

$$f(\hat{\mathbf{x}}) \ge f(\overline{\mathbf{x}}) \ge \hat{f}(\hat{\mathbf{x}}) - c\left(1 + \sum_{i=1}^{m} \hat{u}_i\right) \ge f(\hat{\mathbf{x}}) - c\left(1 + \sum_{i=1}^{m} \hat{u}_i\right).$$

This completes the proof.

In Theorem 11.3.4, the Lagrangian multipliers $\hat{u}_i$ for $i = 1,\dots, m$ 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 $v = 1,\dots, k_j, j \notin L$ be the grid points considered so far. Let $\hat{x}_j$ for $j \in L$ and $\hat{\lambda}_{vj}$ for $v = 1,\dots, k_j, j \notin L$, solve Problem LAP. Furthermore, let $\hat{u}_i \ge 0$ for $i = 1,\dots, m$ 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 $j \notin L$:

$$\begin{aligned} \text{PS: Minimize } & f_j(x_j)+\sum_{i=1}^{m}\hat{u}_i g_{ij}(x_j)+\hat{v}_j \\ \text{subject to } & a_j \leq x_j \leq b_j. \end{aligned}$$

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_{vj}, $v = 1,\ldots, k_j$ 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 $\hat{\lambda}_{vj}$ for $v = 1,\ldots, k_j$, $j \notin L$, 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 $\sum_{v=1}^{k_j}\lambda_{vj} = 1$ in Problem LAP. Now, for each $j \notin L$, consider the following problem:

$$\begin{aligned} \text{Minimize } & f_j(x_j)+\sum_{i=1}^{m}\hat{u}_i g_{ij}(x_j) \\ \text{subject to } & a_j \leq x_j \leq b_j, \end{aligned}$$

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 $\hat{x}_j = \sum_{\nu=1}^{k_j} \hat{\lambda}_{\nu j} x_{\nu j}$ 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

$$\sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} \hat{u}_i \left[\sum_{j=1}^{n} g_{ij}(x_j) - p_i \right] \geq \sum_{j \notin L} \overline{z}_j - \sum_{i=1}^{m} \hat{u}_i p_i \qquad \text{for all } a_j \leq x_j \leq b_j. \tag{11.40}$$

Noting that $a_j \le \overline{x}_j \le b_j$, $\sum_{j=1}^n g_{ij}(\overline{x}_j) \le p_i$, and that $\hat{u}_i \ge 0$, (11.40) implies that $\sum_{j=1}^n f_j(\overline{x}_j) \ge \sum_{j \notin L} \overline{z}_j - \sum_{i=1}^m \hat{u}_i p_i$, 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 $\sum_{j=1}^n f_j(\overline{x}_j) \le \sum_{j=1}^n f_j(\hat{x}_j)$. Finally, by the convexity of f_j for $j \notin L$, we have

$$\begin{aligned}
\sum_{j=1}^{m} f_j(\hat{x}_j) &= \sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} f_j(\hat{x}_j) \\
&= \sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} f_j\left[\sum_{v=1}^{k_j} \hat{\lambda}_{vj} x_{vj}\right] \\
&\le \sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} \sum_{v=1}^{k_j} \hat{\lambda}_{vj} f_j(x_{vj}) \\
&= \hat{z}.
\end{aligned}$$

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

$$f_j(\hat{x}_j) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(\hat{x}_j) = 0 \qquad \text{for } j \in L \tag{11.41}$$

$$\hat{\lambda}_{vj}\left[f_j(x_{vj}) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_{vj}) + \hat{v}_j\right] = 0 \qquad \text{for } v = 1, \ldots k_j, j \notin L \tag{11.42}$$

$$\hat{u}_i\left[\sum_{j \in L} g_{ij}(\hat{x}_j) + \sum_{j \notin L} \sum_{v=1}^{k_j} \hat{\lambda}_{vj} g_{ij}(x_{vj}) - p_i\right] = 0 \qquad \text{for } i = 1, \ldots, m. \tag{11.43}$$

Summing (11.41) over $j \in L$ and (11.42) over $v = 1, \ldots, k_j, j \notin L$, we get

$$\begin{aligned}
&\left[\sum_{j \in L} f_j(\hat{x}_j) + \sum_{j \notin L} \sum_{v=1}^{k_j} \hat{\lambda}_{vj} f_j(x_{vj})\right] + \sum_{i=1}^{m} \hat{u}_i \left[\sum_{j \in L} g_{ij}(\hat{x}_j) + \right. \\
&\qquad \left. + \sum_{j \notin L} \sum_{v=1}^{k_j} \hat{\lambda}_{vj} g_{ij}(x_{vj})\right] + \sum_{j \notin L} \sum_{v=1}^{k_j} \hat{\lambda}_{vj} \hat{v}_j = 0.
\end{aligned} \tag{11.44}$$

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

$$\sum_{j \notin L} \bar{z}_j - \sum_{i=1}^{m} \hat{u}_i p_i \leq \sum_{j=1}^{n} f_j(\bar{x}_j). \tag{11.46}$$

Adding (11.45) to (11.46), we get $\sum_{j \notin L}(\bar{z}_j + \hat{v}_j) + \hat{z} \leq \sum_{j=1}^{n} f_j(\bar{x}_j)$. But by assumption in Part 2, $\bar{z}_j + \hat{v}_j \geq 0$ for $j \notin L$. Hence, $\hat{z} \leq \sum_{j=1}^{n} f_j(\bar{x}_j)$; and, using Part 1 of the theorem, we get, $\hat{z} \leq \sum_{j=1}^{n} f_j(\bar{x}_j) \leq \sum_{j=1}^{n} f_j(\hat{x}_j) \leq \hat{z}$. This implies that $\sum_{j=1}^{n} f_j(\bar{x}_j) = \sum_{j=1}^{n} f_j(\hat{x}_j)$. 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 $\bar{z}_j < -v_j$. 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, $f_j(x_{\gamma j}) + \sum_{i=1}^{m} \hat{u}_i g_{ij}(x_{\gamma j}) + \hat{v}_j \geq 0$, 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 $\sum_{j=1}^{n} g_{ij}(x_j) \leq 0$ for $i = 1, \ldots, m$ and $x_j \geq 0$ 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 $a_j, b_j \geq 0$ 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 $\mathbf{B}^{-1}(2.75, 7.56, 1, 0)^t = (2.75, 7.56, 1, 0)^t$. 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 $\hat{\mathbf{x}} = (1.25, 3.75, 0)^t$ 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 $\{\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ 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 $\mathbf{d}$, 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:

$$\begin{aligned} \text{Minimize} \quad & \frac{-2x_1 + x_2 + 2}{x_1 + 3x_2 + 4} \\ \text{subject to} \quad & -x_1 + x_2 \le 4 \\ & x_2 \le 6 \\ & 2x_1 + x_2 \le 14 \\ & x_1, x_2 \ge 0. \end{aligned}$$

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
$\mathbf{r}$	−10/16	−2/16	0	0	0	—

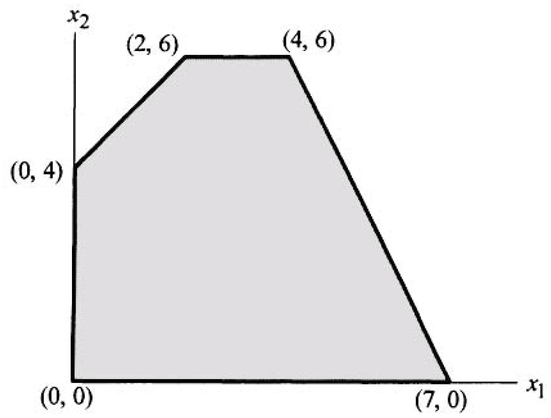


Figure 11.5 Feasible region for Example 11.4.3.

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:

$$\begin{aligned}\text{Minimize}\ & \frac{\mathbf{p}^t\mathbf{x}+\alpha}{\mathbf{q}^t\mathbf{x}+\beta}\\ \text{subject to}\ & \mathbf{Ax}\le\mathbf{b}\\ & \mathbf{x}\ge\mathbf{0}.\end{aligned}$$

Suppose that the set $S = \{\mathbf{x} : \mathbf{Ax} \le \mathbf{b} \text{ and } \mathbf{x} \ge \mathbf{0}\}$ 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 $\mathbf{Ax} \le \mathbf{b}$ by z, the above problem leads to the following linear program:

$$\begin{aligned}\text{Minimize}\ & \mathbf{p}^t\mathbf{y}+\alpha z\\ \text{subject to}\ & \mathbf{Ay}-\mathbf{b}z\le\mathbf{0}\\ & \mathbf{q}^t\mathbf{y}+\beta z=1\\ & \mathbf{y}\ge\mathbf{0}\\ & z\ge 0.\end{aligned}$$

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 $\mathbf{Ay} \le \mathbf{0}$ 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 $\mathbf{Ax} \le \mathbf{b}$ 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

$$\begin{aligned} \text{GP: Minimize } & f(\mathbf{x}) \\ \text{subject to } & g_i(\mathbf{x}) \le 1 \quad \text{for } i = 1,\ldots,m \\ & \mathbf{x} > \mathbf{0}, \end{aligned}$$

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 $\mathbf{a}_k = (a_{k1},\ldots,a_{kn})^t$, 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 $i = 1,\ldots, m$, 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:

$$\text{LD:} \quad \text{Maximize} \quad \{\theta(\mathbf{u}) : \mathbf{u} \geq \mathbf{0}\}, \tag{11.54a}$$

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 $\mathbf{u} \geq 0$, $\theta(\mathbf{u})$ equals that value of $L(\mathbf{y}, \mathbf{u})$ evaluated at the point $\mathbf{y}$ 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 ($\mathbf{y}$) 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 $\mathbf{y}$. However, we would now like to treat $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_M)^t$ as a set of *variables* and write (11.55) in terms of ($\boldsymbol{\delta}$, $\mathbf{u}$) by eliminating $\mathbf{y}$ 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 $\boldsymbol{\delta}$ 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 $\boldsymbol{\delta}$ is orthogonal to each of the n rows of the $n \times M$ matrix $\mathbf{A}$ having columns $\mathbf{a}_1,\ldots,\mathbf{a}_M$.

In the transformed problem, we impose the relationships (11.58) and (11.59) on ($\boldsymbol{\delta}$, $\mathbf{u}$), along with nonnegativity restrictions. However, given any ($\boldsymbol{\delta}$,

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 $i \in \{1,\ldots, m\}$. 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 $u_i = 0$:

$$\text{DGP:} \quad \text{Maximize} \quad \sum_{k=1}^{M} \delta_k \ln\left[\frac{\alpha_k}{\delta_k}\right] + \sum_{i=1}^{m} u_i \ln(u_i) \quad (11.62a)$$

$$\text{subject to} \quad \sum_{k=1}^{M} \delta_k \mathbf{a}_k \equiv \mathbf{A}\boldsymbol{\delta} = \mathbf{0} \quad (11.62b)$$

$$\sum_{k\in J_0} \delta_k = 1 \quad (11.62c)$$

$$\sum_{k\in J_i} \delta_k = u_i \qquad \text{for } i = 1,\ldots, m \quad (11.62d)$$

$$\delta_k \geq 0 \qquad \text{for } k = 1,\ldots, M \quad (11.62e)$$

$$u_i \geq 0 \qquad \text{for } i = 1,\ldots, m. \qquad (11.62f)$$

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 $(\boldsymbol{\delta}^*, \mathbf{u}^*) > \mathbf{0}$ 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

$$\left(\ln\left[\frac{\alpha_k}{\delta_k^*}\right]-1\right)+\mathbf{a}_k^t\mathbf{v}^*+w_i^*=0 \qquad \text{for all } k\in J_i,\ i=0,1,\ldots,m \tag{11.65}$$

$$[\ln(u_i^*)+1]-w_i^*=0 \qquad \text{for all } i=1,\ldots,m, \tag{11.66}$$

where $\mathbf{v}^*$ and $\mathbf{w}^*$ are as defined in the theorem. Substituting for w_i^* from (11.66) into (11.65) for $i = 1,\ldots, m$, we get

$$\mathbf{a}_k^t\mathbf{v}^*=\ln\left[\frac{\delta_k^*}{\alpha_k u_i^*}\right] \qquad \text{for all } k\in J_i,\ i=1,\ldots,m. \tag{11.67}$$

Now consider $\mathbf{y}^*$ as given by (11.64a). We then have from (11.52), (11.53c), (11.62d), and (11.67) that

$$\begin{aligned} G_i(\mathbf{y}^*) &= \sum_{k\in J_i}\tau_k=\sum_{k\in J_i}\alpha_k e^{\mathbf{a}_k^t\mathbf{y}^*} \\ &= \sum_{k\in J_i}\alpha_k\frac{\delta_k^*}{\alpha_k u_i^*}=1 \qquad \text{for } i=1,\ldots,m. \end{aligned} \tag{11.68}$$

Moreover, from (11.52) and (11.65)–(11.67), we have

$$\tau_k=\alpha_k e^{\mathbf{a}_k^t\mathbf{y}^*}=\frac{\delta_k^*}{u_i^*} \quad \text{for all } k\in J_i,\ i=1,\ldots,m. \tag{11.69a}$$

Furthermore, we have

$$\tau_k=\alpha_k e^{\mathbf{a}_k^t\mathbf{y}^*}=\alpha_k e^{[1-w_0^*+\ln(\delta_k^*/\alpha_k)]}=\delta_k^* e^{(1-w_0^*)} \qquad \text{for all } k\in J_0.$$

But $F(\mathbf{y}^*)=\sum_{k\in J_0}\tau_k=e^{(1-w_0^*)}\sum_{k\in J_0}\delta_k^*=e^{1-w_0^*}$ from (11.62c). Hence, we have

$$\tau_k=F(\mathbf{y}^*)\delta_k^* \qquad \text{for all } k\in J_0. \tag{11.69b}$$

Substituting (11.69) into (11.56) and using (11.62b) and (11.68), we get

$$\nabla_{\mathbf{y}} L(\mathbf{y}^*, \mathbf{u}^*) = \sum_{k \in J_0} \delta_k^* \mathbf{a}_k = \sum_{i=1}^{m} \sum_{k \in J_i} \delta_k^* \mathbf{a}_k = \mathbf{0}. \tag{11.70}$$

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 $J_1 = \{3\}$:

$$\begin{aligned} 2\delta_1 + \delta_2 - 2\delta_3 &= 0 \\ \delta_2 - \delta_3 &= 0 \\ \delta_1 + \delta_2 &= 1 \\ \delta_3 &= u_1. \end{aligned}$$

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

$$\begin{aligned} 2y_1 &= \ln\left[\frac{1}{3}\frac{(54\pi V^2)^{1/3}}{2\pi}\right] = \ln\left[\left(\frac{V}{2\pi}\right)^{2/3}\right] \\ y_1 + y_2 &= \ln\left[\frac{2}{3}\frac{(54\pi V^2)^{1/3}}{2\pi}\right] = \ln\left[\left(\frac{2V^2}{\pi^2}\right)^{1/3}\right] \\ -2y_1 - y_2 &= \ln\left[\frac{\pi}{V}\right]. \end{aligned}$$

(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 $V \equiv (256\pi/135)\ \text{cm}^3$.

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 $m = 1$, $n = 2$, $M = 4$, $\text{DD} = M - n - 1 = 1$, $\alpha_1 = 2\pi$, $\alpha_2 = 2\pi$, $\alpha_3 = 2\pi$, $\alpha_4 = V/\pi$, $\mathbf{a}_1^t = (2, 0)$, $\mathbf{a}_2^t = (1, 1)$, $\mathbf{a}_3^t = (0, 1)$, $\mathbf{a}_4^t = (-2, -1)$, 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

$$\delta_1 = (1 - \delta_4),\ \delta_2 = (4\delta_4 - 2),\ \delta_3 = (2 - 3\delta_4), \text{ and } u_1 = \delta_4. \tag{11.72}$$

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 $(\mathbf{w}, \mathbf{z})$ 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 $\mathbf{M}$ 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 $\mathbf{e}$ is a vector of p ones:

$$\begin{aligned}\text{MIP:}\quad \text{Minimize}\quad & \alpha\\ \text{subject to}\quad & \mathbf{0}\leq\mathbf{Mx}+\alpha\mathbf{q}\leq\mathbf{e}-\mathbf{y}\\ & \mathbf{0}\leq\mathbf{x}\leq\mathbf{y}\\ & \mathbf{y}\text{ binary},\ 0\leq\alpha\leq 1.\end{aligned}$$

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 $(\mathbf{w}, \mathbf{z})$ be a complementary basic solution with $\mathbf{z} \geq \mathbf{0}$. If $\mathbf{w} \geq \mathbf{0}$, stop; $(\mathbf{w}, \mathbf{z})$ 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 $\mathbf{A}$ and $\mathbf{B}$, where a_{ij} and b_{ij} are the (i, j)th entries of $\mathbf{A}$ and $\mathbf{B}$, 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 $\mathbf{M}$. 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 $\mathbf{x}_B$, $\mathbf{u}$, and $\mathbf{v}_N$. Note that the solution satisfies all the constraints except possibly $\mathbf{v} \geq \mathbf{0}$. Since $\mathbf{u}$ 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:

$$\begin{aligned} \text{Minimize } & 3x_1^2 + 2x_2^2 - x_1x_2 \\ \text{subject to } & -2x_1 + x_2 \leq 0 \\ & 2x_1 + 3x_2 \geq 6 \\ & 6x_1 + x_2 \leq 12 \\ & x_1, x_2 \geq 0. \end{aligned}$$

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]:

$$\begin{aligned} \text{Minimize } & \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + 3x_1 + 7x_3 + x_4 \\ \text{subject to } & x_1 + 2x_2 + x_3 = 8 \\ & x_1 + 2x_2 + x_4 = 5 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

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 $\mathbf{x}$ 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 $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, 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:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{H}\mathbf{x}+\mathbf{A}^t\mathbf{u}-\mathbf{v} &= -\mathbf{c} \\ \mathbf{x},\mathbf{u} &\geq \mathbf{0} \\ \mathbf{v}^t\mathbf{x} &= 0. \end{aligned}$$

The method first finds a starting basic feasible solution to the system $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$. 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 $\mathbf{A}\mathbf{x} = \mathbf{b}$ is replaced by $\mathbf{A}\mathbf{x} \leq \mathbf{b}$?

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 nonempty.

(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 $\mathbf{A}\mathbf{x} \le \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$, 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 $\mathbf{q}$ 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 $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}$; 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:

$$\text{QP}(\mathbf{x}_k)\text{: Minimize}\left\{(\mathbf{c}+\mathbf{H}\mathbf{x}_k)^t\mathbf{d}+\frac{1}{2}\mathbf{d}^t\mathbf{H}\mathbf{d}: \mathbf{A}_i^t\mathbf{d}=\mathbf{0} \text{ for all } i\in W_k\right\}.$$

Let $\mathbf{d}_k$ 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 $\mathbf{d}_k = \mathbf{0}$, 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:

$$\begin{aligned} \mathbf{A}\mathbf{x}+\mathbf{x}_s &= \mathbf{b} \\ \mathbf{H}\mathbf{x}-\mathbf{u}+\mathbf{A}^t\mathbf{v} &= -\mathbf{c} \\ \mathbf{u}^t\mathbf{x}+\mathbf{v}^t\mathbf{x}_s &= 0 \\ \mathbf{x},\mathbf{x}_s,\mathbf{u},\mathbf{v} &\ge \mathbf{0}. \end{aligned}$$

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:

$$\begin{aligned} \text{Minimize } & \mathbf{y}^t\tilde{\mathbf{y}} \\ \text{subject to } & \mathbf{E}\mathbf{y} = \mathbf{d} \\ & \mathbf{y} \ge \mathbf{0}. \end{aligned}$$

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:

$$\begin{aligned} \text{Minimize } & \tilde{\mathbf{w}}_k^t\mathbf{y} \\ \text{subject to } & \mathbf{E}\mathbf{y} = \mathbf{d} \\ & \mathbf{y} \ge \mathbf{0}. \end{aligned}$$

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:

$$\begin{aligned} \text{Minimize } & -3x_1 - 5x_2 + 2x_1^2 + x_2^2 \\ \text{subject to } & 3x_1 + 2x_2 \le 6 \\ & 4x_1 + x_2 \le 4 \\ & x_1, x_2 \ge 0. \end{aligned}$$

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:

$$\begin{aligned} \text{Maximize } & z \\ \text{subject to } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{H}\mathbf{x} + \mathbf{A}^t\mathbf{u} - \mathbf{v} + z\mathbf{c} = \mathbf{0} \\ & \mathbf{x} \ge \mathbf{0}, \mathbf{v} \ge \mathbf{0}, z \ge 0, \qquad \mathbf{u} \text{ unrestricted.} \end{aligned}$$

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.

$$\text{LP: Minimize } \{-w_{11} - w_{22} + 24x_1 - 144 : [(24-x_1)s_1]_L \geq 0, [x_1, s_2]_L \geq 0, [(24-x_1)x_1]_L \geq 0, [(24-x_1)x_1s_1]_L \geq 0, [(24-x_1)x_2s_1]_L \geq 0, [(24-x_1)x_1s_2]_L \geq 0, \text{ and } [x_1x_2s_2]_L \geq 0.$$

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 $\mathbf{W}$ denote the $n \times n$ symmetric matrix having $w_{(ij)}$ as its (i, j)th element.

a. Verify that $\mathbf{W}$ 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 $\mathbf{X}$ being symmetric) that $\boldsymbol{\alpha}^t\mathbf{X}\boldsymbol{\alpha} \geq 0$, $\forall\ \boldsymbol{\alpha} \in R^{n+1}$, such that $\|\boldsymbol{\alpha}\| = 1$.
d. Suppose now that we solve the LP(Ω) relaxation (11.17) and obtain a solution $(\overline{\mathbf{x}}, \overline{\mathbf{w}})$, and we use this solution to compose the matrix $\overline{\mathbf{X}}$. 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 $\bar{\alpha} \in R^{n+1}$ of the type in Part c for which $\bar{\alpha}^t \overline{\mathbf{X}} \bar{\alpha} < 0$. Hence, show that it is valid to impose the *semidefinite cut* $\bar{\alpha}^t \overline{\mathbf{X}} \bar{\alpha} \geq 0$ 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*

$$\text{PP: Minimize}\{\phi_0(\mathbf{x}) : \phi_r(\mathbf{x}) \geq \beta_r,\ r = 1,\ldots,m,\ \mathbf{x} \in \Omega\},$$

where $\Omega = \{\mathbf{x} : 0 \leq \ell_j \leq x_j \leq u_j < \infty \text{ for } j = 1,\ldots, n\}$ and where $\phi_r(\mathbf{x}) \equiv \sum_{t \in T_r} \alpha_{rt} \prod_{j \in J_{rt}} x_j$ for $r = 0, 1,\ldots, m$. Hence, for each $r = 0, 1,\ldots, m$, 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 $N \equiv \{1,\ldots, n\}$. (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 $\overline{N} \equiv \{N,\ldots, N\}$ represent δ copies of N. Hence, each $J_{rt} \subseteq \overline{N}$ with $1 \leq |J_{rt}| \leq \delta$, $\forall\ t \in T_r$, $r = 0, 1,\ldots, R$.

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): Minimize$\{[\phi_0(\mathbf{x})]_L : [\phi_r(\mathbf{x})]_L \geq \beta_r,\ r = 1,\ldots, R,\ x \in \Omega,$ plus the inequalities (11.75)$\}$. (11.77)

Show that $v[\text{LP}(\Omega)] \leq v[\text{PP}]$. Moreover, show that if $(\overline{\mathbf{x}}, \overline{\mathbf{X}})$ solves $\text{LP}(\Omega)$ and satisfies (11.76), then $\overline{\mathbf{x}}$ solves PP.

c. Let $(\overline{\mathbf{x}}, \overline{\mathbf{X}})$ be any feasible solution to $\text{LP}(\Omega)$. Suppose that $\overline{x}_p = \ell_p$ or $\overline{x}_p = u_p$ for some $p \in \{1,..., n\}$. Then show that

$$\overline{X}_{(J \cup \{p\})} = \overline{x}_p \overline{X}_J, \ \forall \ J \subseteq \overline{N}, \ 1 \leq |J| \leq \delta - 1$$

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 $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, 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 $V(\mathbf{x}_S) \neq \varnothing$, 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 $V(\mathbf{x}_{S_1}) \neq \varnothing$ 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:

$$\begin{aligned}&\text{Minimize} \quad \frac{1}{2x_1+1} + 3x_2^3\\ &\text{subject to} \quad 2x_1^2 - x_2^3 \le 4\\ &\qquad\qquad\qquad x_1, x_2 \ge 6.\end{aligned}$$

[11.31] Consider the following problem:

$$\begin{aligned}&\text{Minimize} \quad -2x_1 + 3x_2 + 3x_1^2 - 2x_1x_2 + 2x_2^2\\ &\text{subject to} \quad 2x_1 + x_2 \le 6\\ &\qquad x_1^2 + x_2^2 = 9\\ &\qquad x_1, x_2 \ge 0.\end{aligned}$$

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:

$$\begin{aligned}&\text{Minimize} \quad 3e^{2x_1} + 2x_1^2 + 3x_1 + 2x_2^2 - 5x_2 + 3x_3\\ &\text{subject to} \quad x_1^2 + e^{x_2} + 6x_3 \le 15\\ &\qquad x_1^4 - x_2 + 5x_3 \le 25\\ &\qquad 0 \le x_1 \le 4\\ &\qquad 0 \le x_2 \le 2\\ &\qquad 0 \le x_3 \le \infty.\end{aligned}$$

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 $[\mu_v, \mu_{v+1}]$. Then x can be represented as $x = \mu_1 + \sum_{i=1}^{v} \delta_i \Delta_i$, and $\theta(x)$ can be approximated by $\hat{\theta}(x) = \theta_1 + \sum_{i=1}^{v} \delta_i \Delta\theta_i$, where $\delta_v \in [0, 1]$, $\delta_i = 1$ for $i = 1, \ldots, v-1$, 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_{vj} for $v = 1, \ldots, k_j + 1$ 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 $\lambda = \mathbf{T}\delta$, where $\mathbf{T}$ 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 $\{(x_1,x_2): 3x_1-x_2+2\neq 0\}$? 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

$$\begin{aligned} \text{Minimize } & 35x_1^2x_2 + 15x_2x_3 \\ \text{subject to } & \frac{2}{5}x_1^{-1}x_2^{-1/3} + \frac{3}{5}x_2^{-2}x_3^{-4/3} \le 1 \\ & \mathbf{x} > \mathbf{0}. \end{aligned}$$

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 $2x_1^{-1/3}x_2^{1/6} + [(3/5)x_1^{1/2}x_2^{3/4} + (2/5)x_1^{2/3}x_2]^{1/2} x_1^{3/4}x_2^{-1/3}$.

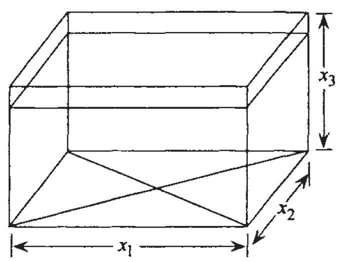
[11.48] Consider the geometric programming problem to

$$\begin{aligned} \text{Minimize } & 25x_1^{-2}x_2^{-1/2}x_3^{-1} + 20x_1^2x_3 + 30x_1x_2^2x_3 \\ \text{subject to } & \frac{5}{3}x_1^{-1}x_2^{-2} + \frac{4}{3}x_2^{1/2}x_3^{-2} \le 1 \\ & \mathbf{x} > \mathbf{0}. \end{aligned}$$

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 $\mathbf{M}$ 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 $\mathbf{M}$ is positive semidefinite, or if $\mathbf{M}$ 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 $\mathbf{M}$ 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 $\mathbf{M}$ 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 $\mathbf{M}$ 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].