PRACTICAL BILEVEL OPTIMIZATION

Algorithms and Applications

Jonathan F. Bard

The University of Texas
Graduate Program in Operations Research
Department of Mechanical Engineering
Austin, Texas
USA

KLUWER ACADEMIC PUBLISHERS
Boston/London/Dordrecht

CONTENTS

2 LINEAR PROGRAMMING 17

2.1 Introduction 17

2.1.1 Basic Solutions 22

2.1.2 Fundamental Theorem 23

2.1.3 Convex Properties 25

2.2 Simplex Method 29

2.2.1 Pivoting 30

2.2.2 Determining the Leaving Variable 32

2.2.3 Moving toward Optimality 33

2.2.4 Degeneracy and Cycling 37

2.3 Geometry of Simplex Method 41

2.3.1 Finiteness of Algorithm 42

2.3.2 Adjacency and Bounded Edges 42

2.3.3 Unboundedness 44

2.3.4 Finding Adjacent Extreme Points 47

2.3.5 Main Geometric Argument 47

2.3.6 Alternative Optima and Uniqueness 48

2.3.7 Ranking Extreme Points 49

2.4 Additional Features 51

2.4.1 Phase 1 and Artificial Variables 51

2.4.2 Bounded Variables 54

2.4.3 Kuhn-Tucker Conditions and the Linear Complementarity Problem 57

2.5 Duality 59

2.5.1 Primal–Dual Relationship 61

2.5.2 Dual Theorems 61

2.5.3 Economic Interpretation 66

2.5.4 Sensitivity Analysis 66

2.5.5 Dual Simplex Method 72

REFERENCES 76

2

LINEAR PROGRAMMING

2.1 INTRODUCTION

In general, constrained optimization problems can be written as

$$\min\{f(\boldsymbol{x}) \ : \ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$$

where $\boldsymbol{x} \in R^n$, $f : R^n \rightarrow R^1$, $\boldsymbol{h} : R^n \rightarrow R^m$ and $\boldsymbol{g} : R^n \rightarrow R^q$. The simplest form of this problem is realized when the functions $f(\boldsymbol{x})$, $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$ are all linear in $\boldsymbol{x}$. The resulting model is known as a linear program (LP) and plays a central role in virtually every branch of optimization. Many real situations can be formulated or approximated as LPs, optimal solutions are relatively easy to calculate, and computer codes for solving very large instances consisting of millions of variables and tens of thousands of constraints are commercially available. Another attractive feature of linear programs is that various subsidiary questions related, for example, to the sensitivity of the optimal solution to changes in the data and the inclusion of additional variables and constraints can be analyzed with little effort.

The importance of this topic cannot be overstated. Linear programming algorithms are frequently used as subroutines for solving more difficult optimization problems in nonlinear and integer programming. The linear bilevel programming problem (BLPP) or Stackelberg game discussed in Chapter **??** is a case in point. The primary purpose of this chapter is to provide a foundation for the presentation of algorithms for solving that problem. The most successful algorithms in this regard have tried to exploit the polyhedral nature of the feasible region of the BLPP using branch and bound techniques to narrow the solution space. This can be done most effectively with some form of vertex enumeration — the idea behind the simplex method of linear programming. Therefore, we concentrate on this method alone with only passing reference to the more recently developed techniques that restrict the search for a solution to the interior of the feasible region. These are known as *interior point methods* (e.g., see [C4, H9, K2, L8, M7, M8]).

The linear programming problem was initially stated using quasi-economic terminology which to some extent obscures the basic numerical processes that are involved in the analysis. Our presentation aims to make these processes clear while retaining the traditional nomenclature. One main feature of the traditional approach is that the linear programming is expressed in *standard form*

$$\min \ f(\boldsymbol{x}) \triangleq \boldsymbol{c}\boldsymbol{x} \tag{2.1a}$$

$$\text{subject to } \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.1b}$$

$$\boldsymbol{x} \geqq \boldsymbol{0} \tag{2.1c}$$

where $\boldsymbol{c}$ is an n-dimensional row vector, $\boldsymbol{b}$ an m-dimensional column vector, $\boldsymbol{A}$ an $m \times n$ matrix with $m \leqq n$, and $\boldsymbol{x} \in R^n$. Thus the allowable constraints on the variables are either linear equations or nonnegative bounds. The coefficients $\boldsymbol{c}$ in the objective function are referred to as costs, the $\boldsymbol{A}$ matrix comprises the technological coefficients, and the right-hand side vector $\boldsymbol{b}$ denotes the resource levels. It is generally assumed that the data $(\boldsymbol{c}, \boldsymbol{A}, \boldsymbol{b})$ defining the problem are known and deterministic. An example with four variables ($n = 4$) and two technological constraints ($m = 2$) is

$$\begin{array}{rl} \min & x_1 + 2x_2 + 3x_3 + 4x_4 \\ \text{subject to} & \begin{array}{rcl} x_1 + x_2 + x_3 + x_4 & = & 1 \\ x_1 + x_3 - 3x_4 & = & \frac{1}{2} \end{array} \\ & x_1, x_2, x_3, x_4 \geqq 0 \end{array} \tag{2.2}$$

More general LPs can be reduced to standard form without undue difficulty, albeit with some possible loss in efficiency. For instance, a general linear inequality $\boldsymbol{a}^T\boldsymbol{x} \leq b$ can be transformed by adding a *slack variable* $x_s \in R^1$ to the left-hand side to get $\boldsymbol{a}^T\boldsymbol{x} + x_s = b$ with the nonnegativity requirement $x_s \geqq 0$. To convert a "greater than or equal to" inequality of the form $\boldsymbol{a}^T\boldsymbol{x} \geqq b$ to an equation, a nonnegative *surplus variable* x_s is subtracted from the left-hand side to get $\boldsymbol{a}^T\boldsymbol{x} - x_s = b$.

Standard form usually assumes that the resource vector $\boldsymbol{b}$ is greater than or equal to zero. It should be clear that by suitably multiplying by minus one and adjoining slack and surplus variables, any set of linear inequalities can be converted to the form given in (2.1b) with $\boldsymbol{b} \geqq \boldsymbol{0}$. More general bounds $x_j \geqq l_j$ can be dealt with by a shift in the origin; i.e., by replacing x_j with $\widehat{x}_j + l_j$, where $\widehat{x}_j \geqq 0$. In fact very little is lost in complexity if the bounds in (2.1c) are expressed as $\boldsymbol{l} \leqq \boldsymbol{x} \leqq \boldsymbol{u}$. Also, if a linear program is given in standard form except that one or more of the decision variables is not required to be nonnegative, the problem can be transformed into standard form by either of two simple techniques.

Free variable method 1: Suppose in (2.1c), x_j is not restricted to be nonnegative and hence can take on any real value. We then write $x_j = x_j^+ - x_j^-$ and require that $x_j^+ \geqq 0$ and $x_j^- \geqq 0$. If we substitute $x_j^+ - x_j^-$ for x_j everywhere in (2.1a) and (2.1b), the linearity of the constraints is preserved and all variables are now required to be nonnegative. The problem is then expressed in terms of the $n + 1$ variables

$x_1, x_2, \ldots, x_j^+, x_j^-, \ldots, x_n$. Of course, there is a certain amount of redundancy introduced by this technique because a constant added to x_j^+ and x_j^- does not change x_j (that is, the representation of a given value of x_j is not unique). Nevertheless, this does not hinder the simplex method.

Free variables method 2: A second approach when x_j is unrestricted in sign is to eliminate x_j together with one of the constraint equations. Take any one of the m equations in (2.1b) that has a nonzero coefficient for x_j, say, equation 1,

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1j}x_j + \cdots + a_{1n}x_n = b_1 \tag{2.3}$$

where $a_{1j} \neq 0$. The variable x_j can be expressed as a linear combination of the other variables plus a constant. Substituting this expression for x_j everywhere in (2.1a) and (2.1b) gives a new problem of exactly the same form but in terms of the variables $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ only. Moreover, the equation used to determine x_j is now identically zero and can also be eliminated. This scheme is valid since any combination of nonnegative variables $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ produces a feasible x_j from (2.3). Note that this would not necessarily be true if x_j were originally restricted to be nonnegative. As a result of this simplification, we obtain a standard linear program having $n-1$ variables and $m-1$ equality constraints. After a solution is found to the reduced problem the value of x_j can be determined from (2.3).

As is true with any optimization problem, it is important to realize that a linear program may have no solution, either because there is no feasible point (the problem is *infeasible*), or because $f(x) \to -\infty$ for $\boldsymbol{x}$ in the feasible region (the problem is *unbounded*). Nevertheless, we show that there is no difficulty in detecting these situations so we concentrate on the usual case in which a (possibly nonunique) solution exists. It is also convenient to assume that the equality constraints (2.1b) are linearly independent, implying that the rank of the $\boldsymbol{A}$ matrix is m. In theory, this can always be achieved by either removing dependent equations or adding artificial variables, although in practice numerical difficulties might arise if this dependence is not detected.

Considering (2.1) in more detail, if $m = n$, then the equations $\boldsymbol{Ax} = \boldsymbol{b}$ determine a unique solution under the independence assumption, and the objective function $\boldsymbol{cx}$ and the bounds $\boldsymbol{x} \geqq \boldsymbol{0}$ play no part. In most cases, however, $m < n$ so that the system $\boldsymbol{Ax} = \boldsymbol{b}$ is underdetermined and $n - m$ degrees of freedom remain. In particular, the system can determine only m variables, given values for the remaining $n-m$ variables. For example, the equations $\boldsymbol{Ax} = \boldsymbol{b}$ in (2.2) can be rearranged as

$$\begin{aligned} x_1 &= \tfrac{1}{2} - x_3 + 3x_4 \\ x_2 &= \tfrac{1}{2} \qquad - 4x_4 \end{aligned} \tag{2.4}$$

which determines x_1 and x_2 in terms of x_3 and x_4, or alternatively as

$$\begin{aligned} x_1 &= \tfrac{7}{8} - \tfrac{3}{4}x_2 - x_3 \\ x_4 &= \tfrac{1}{8} - \tfrac{1}{4}x_2 \end{aligned} \tag{2.5}$$

which determines x_1 and x_4 from x_2 and x_3, and so on. It is important to consider what values these remaining $n - m$ variables can take in the standard form of the problem. The objective function $\boldsymbol{cx}$ is linear and so contains no curvature which can give rise to a minimizing point. Hence such a point must be created by the conditions $x_j \geqq 0$ becoming active on the boundary of the feasible region. For example, if (2.5) is used to eliminate the variable x_1 and x_4 from problem (2.2), then the objective function can be written as

$$f = x_1 + 2x_2 + 3x_3 + 4x_4 = \frac{11}{8} + \frac{1}{4}x_2 + 2x_3 \tag{2.6}$$

which clearly has no minimum value unless the conditions $x_2 \geqq 0$, $x_3 \geqq 0$ are imposed. In this case, the minimum occurs when $x_2 = 0$, $x_3 = 0$.

To illustrate the nature of an LP solution graphically, consider the simpler constraint set $2x_1 + x_2 = 3$ and $x_1 \geqq 0$, $x_2 \geqq 0$. The feasible region is shown in Fig. 2.1 as the bold line joining points $\boldsymbol{a} = (0, 3)$ and $\boldsymbol{b} = (\frac{3}{2}, 0)$. When the objective function $f(\boldsymbol{x})$ is linear the solution must occur at either $\boldsymbol{a}$ or $\boldsymbol{b}$ with either $x_1 = 0$ or $x_2 = 0$. For example, for $f = -3x_1 + 4x_2$ or $f = x_1 + x_2$ the solution occurs at $\boldsymbol{a}$ (try other linear functions). If $f = 2x_1 + x_2$, however, any point on the line segment connecting $\boldsymbol{a}$ and $\boldsymbol{b}$ provides the same objective function value, implying that a solution need not be unique.

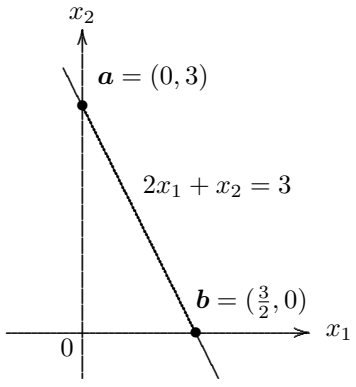

Figure 2.1 Constraints for simple LP

The example in Fig. 2.1 demonstrates that if the feasible region is bounded, a solution of an LP problem in standard form always exists at one particular extreme point or vertex of the feasible region, with at least $n - m$ variables equal to zero and the remaining m variables being uniquely determined by the equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and taking nonnegative values. This result is fundamental to the development of algorithms for solving LPs and will be established vigorously in Section 2.1.3. Recall Weierstrass' Theorem which states in its most elementary form that a continuous function f defined on a compact set S has a minimum point in S (see Section **??**). If S is not compact, it still can be shown that if a finite solution exists to an LP, there is at least one vertex solution. In our case the set $S = \{\boldsymbol{x} \in R^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqq 0\}$ denotes the feasible region (2.1b)–(2.1c).

The main difficulty in linear programming is to find which $n - m$ variables take zero value at the solution. The brute force approach to making this determination is to enumerate all $\binom{n}{m}$ combinations of variables, solve the corresponding set of m linear equations in m variables, evaluate the objective function at each feasible combination, and select the best result. For all but the smallest problems, this approach is highly inefficient. For the linear BLPP, a variant of this idea has proven successful. In Section 2.3.7, we give an algorithm for ranking (and hence enumerating) all vertices of a polytope.

The earliest algorithm for solving the linear program is due to George Dantzig and is called the simplex method. The basic idea is to traverse the vertices of the underlying polytope in a systematic way that avoids investigating infeasible points. Different combinations of variables are examined one at a time without ever returning to a previously explored combination (vertex). The simplex method still predominates today with different variations existing, depending on which intermediate quantities are computed and which techniques are used to handle the linear algebra. The earliest tableau form was quickly superseded by the more efficient revised simplex method that is described in Section 2.2. In the last few years, computational schemes using matrix factorizations have been adopted to control round-off errors more effectively. For large sparse problems with up to 10^6 variables, LU decomposition with threshold pivoting (see [S17]) is the method of choice for storing a representation of the basis inverse. This replaced the product form of the inverse which can still be found in some older implementations. Larger LPs that have a network structure can be solved efficiently with specialized codes. Nevertheless, even if a problem has a well-defined solution, all current simplex-type methods may have difficulty finding it if degeneracy is present. This issue is taken up in Section 2.2.4.

2.1.1 Basic Solutions

In working toward a solution to a linear program, it is convenient to start with an analysis of the system of equations $\boldsymbol{Ax} = \boldsymbol{b}$ given in (2.1b), where once again, $\boldsymbol{x}$ is an n-dimensional vector, $\boldsymbol{b}$ is an m-dimensional vector, and $\boldsymbol{A}$ is an $m \times n$ matrix. Suppose that from the n columns of $\boldsymbol{A}$ we select a set of m linearly independent columns (such a set exists if the rank of $\boldsymbol{A}$ is m). For notational simplicity assume that we select the first m columns of $\boldsymbol{A}$ and denote the corresponding $m \times m$ matrix by $\boldsymbol{B}$. The matrix $\boldsymbol{B}$ is then nonsingular and can be uniquely determined by solving the equation

$$\boldsymbol{Bx}_B = \boldsymbol{b}$$

for the m-dimensional vector $\boldsymbol{x}_B$. By putting $\boldsymbol{x} = (\boldsymbol{x}_B, \boldsymbol{0})$, that is, by setting the first m components of $\boldsymbol{x}$ to those of $\boldsymbol{x}_B$ and the remaining components to zero, we obtain a solution to $\boldsymbol{Ax} = \boldsymbol{b}$. This leads to the following definition.

Definition 2.1.1 Given a set of m simultaneous linear equations (2.1b) in n unknowns, let $\boldsymbol{B}$ be any nonsingular $m \times m$ matrix made up of columns of $\boldsymbol{A}$. If all the $n - m$ components of $\boldsymbol{x}$ not associated with columns of $\boldsymbol{B}$ are set equal to zero, the solution to the resulting set of equations is said to be a *basic solution* to (2.1b) with respect to the basis $\boldsymbol{B}$. The components of $\boldsymbol{x}$ associated with columns of $\boldsymbol{B}$ are called *basic variables*.

The m linearly independent columns of $\boldsymbol{B}$ can be regarded as a basis for the space R^m, hence the terminology. A basic solution corresponds to an expression for the vector $\boldsymbol{b}$ as a linear combination of these basis vectors. This interpretation is further discussed in Section 2.1.2.

In some instances, (2.1b) may have no basic solution. However, to avoid trivialities and nonessential difficulties, a number of elementary assumptions regarding the nature of $\boldsymbol{A}$ will be made. These have already been mentioned: the first is that the number of x variables exceeds the number of equality constraints ($n > m$); the second is that the rows of $\boldsymbol{A}$ are linearly independent. A linear dependence among the rows of $\boldsymbol{A}$ would imply either a redundancy in the m equations that could be eliminated or contradictory constraints and hence no solution to (2.1b).

Given the assumption that $\boldsymbol{A}$ has full row rank, the system $\boldsymbol{Ax} = \boldsymbol{b}$ always has a solution and, in fact, it will always have one basic solution; however, the basic variables in a solution are not necessarily all nonzero. This is noted in the following definition.

Definition 2.1.2 A *degenerate basic solution* is said to occur if one or more of the basic variables in a basic solution has value zero.

Thus in a nondegenerate basic solution the basic variables, and hence the basis $\boldsymbol{B}$, can be immediately identified from the positive components of the solution. The same

cannot be said for a degenerate basic solution because a subset of the zero-valued basic and nonbasic variables can be interchanged. This implies some amount of ambiguity but does not cause any difficulties.

So far in the discussion we have only treated the equality constraints of the linear program to the exclusion of the nonnegativity constraints on the variables. We now want to consider the full set of constraints given by (2.1b) and (2.1c) for an LP in standard form.

Definition 2.1.3 A vector $\boldsymbol{x} \in S = \{\boldsymbol{x} \in R^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geqq \boldsymbol{0}\}$ is said to be *feasible* to the linear programming problem in standard form; a feasible solution that is also basic is said to be a *basic feasible solution.* If this solution is degenerate, it is called a *degenerate basic feasible solution.*

2.1.2 Fundamental Theorem

In this section, we establish the relationship between optimality and basic feasible solutions in the fundamental theorem of linear programming. The proof is as important as the theorem itself because it underlies the development of the simplex algorithm. The results tell us that when seeking a solution to an LP it is only necessary to consider basic feasible solutions.

Theorem 2.1.1 Given a linear program in standard form (2.1) where $\boldsymbol{A}$ is an $m \times n$ matrix of rank m,

i) if there is a feasible solution, there is a basic feasible solution;

ii) if there is an optimal feasible solution, there is an optimal basic feasible solution.

Proof of (i): Denote the columns of $\boldsymbol{A}$ by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a feasible solution. In terms of the columns of $\boldsymbol{A}$, a solution can be written as

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \cdots + \boldsymbol{a}_n x_n = \boldsymbol{b}$$

Assume that p of the variables x_j are greater than zero, and for convenience, that they are the first p variables. This gives

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \cdots + \boldsymbol{a}_p x_p = \boldsymbol{b} \tag{2.7}$$

Two cases must now be considered corresponding to whether or not the columns $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, ..., $\boldsymbol{a}_p$ are linearly independent.

Case 1: Assume $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ are linearly independent, implying that $p \leqq m$. If $p = m$, the solution is basic and the proof is complete. If $p < m$, the fact that $\boldsymbol{A}$ has rank m means that $m - p$ vectors can be found so that the resulting set of m vectors is linearly independent (see Corollary 2.1.1). Assigning the value zero to the corresponding $m - p$ variables yields a (degenerate) basic feasible solution.

Case 2: Assume $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ are linearly dependent, implying that there is a non-trivial linear combination of these vectors that is zero. Thus there are constants $w_1, w_2, \ldots, w_p$ at least one of which is positive such that

$$\boldsymbol{a}_1 w_1 + \boldsymbol{a}_2 w_2 + \cdots + \boldsymbol{a}_p w_p = \boldsymbol{0} \tag{2.8}$$

Multiplying each equation in (2.8) by a scalar ε and subtracting it from (2.7), we obtain

$$\boldsymbol{a}_1(x_1 - \varepsilon w_1) + \boldsymbol{a}_2(x_2 - \varepsilon w_2) + \cdots + \boldsymbol{a}_p(x_p - \varepsilon w_p) = \boldsymbol{b}$$

which holds for every ε. In other words, for each ε the components $(x_j - \varepsilon w_j)$ correspond to a solution of the linear equalities; however, they may violate $(x_j - \varepsilon w_j) \geqq 0$. Denoting $\boldsymbol{w} = (w_1, w_2, \ldots, w_p, 0, 0, \ldots, 0)$, we see that for any ε,

$$\boldsymbol{x} - \varepsilon \boldsymbol{w} \tag{2.9}$$

is a solution to the equality constraints. For $\varepsilon = 0$, this reduces to the original feasible solution $\boldsymbol{x}$. As ε is increased from zero, the various components increase, decrease, or remain the same, depending on whether the corresponding component w_j is positive, negative or zero. Because at least one w_j is positive, at least one component will decrease as ε is increased. Now, let us increase ε to the first point where one or more components become zero; i.e.,

$$\varepsilon = \min \left\{ \frac{x_j}{w_j} \;\middle|\; w_j > 0 \right\}$$

For this value of ε the solution given by (2.9) is feasible and has at most $p-1$ positive variables. Repeating this process as necessary, we can eliminate positive variables until we have a solution with corresponding columns that are linearly independent. The resultant situation reduces to case 1. ∎

Proof of (ii): Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be an optimal feasible solution, and, as in the proof of (i), suppose there are exactly p positive variables $x_1, x_2, \ldots, x_p$. Again there are two cases with case 1 corresponding to linear independence exactly as before.

Case 2 is also the same except that we now must show that for any ε the solution (2.9) is optimal. To see this, note that the value of the solution is given by

$$\boldsymbol{c}\boldsymbol{x} - \varepsilon \boldsymbol{c}\boldsymbol{w} \tag{2.10}$$

For sufficiently small ε positive or negative, $\boldsymbol{x} - \varepsilon \boldsymbol{w}$ is a feasible solution to the LP. If $\boldsymbol{c}\boldsymbol{w} \neq 0$, an ε of small magnitude and proper sign could be determined to render (2.10) smaller than $\boldsymbol{c}\boldsymbol{x}$ while maintaining feasibility. This would violate the assumption that $\boldsymbol{x}$ is optimal, implying $\boldsymbol{c}\boldsymbol{w} = 0$. This establishes that the new feasible solution with fewer positive components is also optimal. The remainder of the proof can be completed exactly as in part (i). ∎

This theorem reduces the task of solving LPs to that of searching over basic feasible solutions. For a problem having n variables and m constraints, there are at most

$n!/m!(n-m)!$ basic solutions corresponding to the number of ways of selecting m of n columns. Although it would be extremely inefficient to examine each combination, by expanding on the technique used to prove Theorem 2.1.1, the simplex algorithm can be derived. Before getting to the specifics, we note that the proof given above is of a simple algebraic character. In the next section the geometric interpretation of the theorem is explored in terms of the general theory of convex sets.

2.1.3 Convex Properties

Thus far in the development of linear programming, we have focused the discussion on common properties of systems of linear equations. A second approach, leading to an alternative derivation of the fundamental theorem and perhaps a clearer geometric understanding of the result, can be pursued in terms of the theory of convex sets. The principal link between the algebraic and geometric theories is the formal relation between basic feasible solutions of linear equalities in standard form and extreme points of polytopes.

Definition 2.1.4 A point $\boldsymbol{x}$ in the convex set $C \subset R^n$ is said to be an *extreme point* of C if there are no two distinct points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ in C such that $\boldsymbol{x} = \alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2$ for some $\alpha \in (0,1)$.

An extreme point is thus a point that does not lie strictly within the line segment connecting two other points in the set. The extreme points of a triangle, for example, are its three vertices; every point on the circumference of a circle is an extreme point.

Theorem 2.1.2 The set of all feasible solutions to the linear programming problem is a convex set.

Proof: For the trivial case where the feasible region S is a singleton, the theorem is of course true. For the more general case, we need to show that every convex combination of any two feasible solutions is also feasible. Assume that there are at least two solutions $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ with

$$\boldsymbol{A}\boldsymbol{x}^1 = \boldsymbol{b},\ \boldsymbol{x}^1 \geqq \boldsymbol{0} \ \text{ and } \ \boldsymbol{A}\boldsymbol{x}^2 = \boldsymbol{b},\ \boldsymbol{x}^2 \geqq \boldsymbol{0}$$

For $0 \leqq \alpha \leqq 1$, let $\boldsymbol{x} = \alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2$ be any convex combination of $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$. We note that all elements of the vector $\boldsymbol{x}$ are nonnegative; i.e., $\boldsymbol{x} \geqq \boldsymbol{0}$. Substituting for $\boldsymbol{x}$ in the linear equalities gives

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}[\alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2] = \alpha A\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{A}\boldsymbol{x}^2 = \alpha\boldsymbol{b} + (1-\alpha)\boldsymbol{b} = \boldsymbol{b}$$

which shows that $\boldsymbol{x}$ is feasible. ∎

As before, we shall denote the convex set of solutions to the linear programming problem by S. Because S is determined by the intersection of a finite number of linear

equalities (2.1b) and inequalities (2.1c), its boundary (assuming S is nonempty) will consist of sections of the corresponding hyperplanes. If S is bounded and nonempty, it is called a convex polyhedron and the LP will have a finite solution. If S is unbounded, it is referred to more generally as a polytope, and the LP may or may not have a finite solution. By Theorem 2.1.2, if a problem has more than one solution, in reality, it has an infinite number of solutions.

Before proceeding with the developments, we note that if S is a *convex polyhedron* then it is equivalently the convex hull of its extreme points. That is, every feasible solution in S can be represented as a convex combination of the extreme feasible solutions in S. This has implications for solving large–scale linear models. (By definition, a convex polyhedron has a finite number of extreme points.) An unbounded S also has a finite number of extreme points, but not all points in S can be represented as a convex combination of these extreme points. More will be said about this in Section 2.3.3. For the moment, we will assume that S is nonempty and bounded. As will be shown in later sections of this chapter, computational procedures are available to determine whether S is empty or whether the problem has an unbounded minimum.

From the above discussion, one might surmise that extreme points play an important role in solving linear programs. We prove this in the following theorem.

Theorem 2.1.3 The objective function (2.1a) assumes its minimum at an extreme point of the convex polyhedron S generated by the set of feasible solutions to the linear program. If it assumes its minimum at more than one extreme point, then it takes on the same value for every convex combination of those particular points.

Proof: The assumption that S is a convex polyhedron means that it has a finite number of extreme points. Let us denote the objective function by $f(\boldsymbol{x})$, the extreme points by $\widehat{\boldsymbol{x}}^1, \widehat{\boldsymbol{x}}^2, \ldots, \widehat{\boldsymbol{x}}^s$, and the optimal solution by $\boldsymbol{x}^*$. This means that $f(\boldsymbol{x}^*) \leqq f(\boldsymbol{x})$ for $\boldsymbol{x} \in S$. If $\boldsymbol{x}^*$ is an extreme point the first part of the theorem is true; if $\boldsymbol{x}^*$ is not an extreme point then we can write $\boldsymbol{x}^*$ as a convex combination of the s extreme points in S: $\boldsymbol{x}^* = \sum_{i=1}^{s} \alpha_i \widehat{\boldsymbol{x}}^i$, for $\alpha_i \geqq 0$ and $\sum_{i=1}^{s} \alpha_i = 1$. Then, noting that $f(\boldsymbol{x})$ is a linear functional, we have

$$
\begin{aligned}
f(\boldsymbol{x}^*) = f\left(\sum_{i=1}^{s} \alpha_i \widehat{\boldsymbol{x}}^i\right) &= f(\alpha_1 \widehat{\boldsymbol{x}}^1 + \alpha_2 \widehat{\boldsymbol{x}}^2 + \cdots + \alpha_s \widehat{\boldsymbol{x}}^s) \\
&= \alpha_1 f(\widehat{\boldsymbol{x}}^1) + \alpha_2 f(\widehat{\boldsymbol{x}}^2) + \cdots + \alpha_s f(\widehat{\boldsymbol{x}}^s) \\
&\geqq \alpha_1 f(\widehat{\boldsymbol{x}}^m) + \alpha_2 f(\widehat{\boldsymbol{x}}^m) + \cdots + \alpha_s f(\widehat{\boldsymbol{x}}^m) \\
&= f(\widehat{\boldsymbol{x}}^m) = f^*
\end{aligned}
$$

where $f(\widehat{\boldsymbol{x}}^m) = \min\{f(\widehat{\boldsymbol{x}}^i) : i = 1, \ldots, s\}$. Because we assumed $f(\boldsymbol{x}^*) \leqq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in S$, we must have $f(\boldsymbol{x}^*) = f(\widehat{\boldsymbol{x}}^m) = f^*$. Therefore, there is an extreme point $\widehat{\boldsymbol{x}}^m$ at which the objective function assumes its minimum value.

To prove the second part of the theorem, let $f(\boldsymbol{x})$ assume its minimum at more than one extreme point, say at $\widehat{\boldsymbol{x}}^1$, $\widehat{\boldsymbol{x}}^2$, ..., $\widehat{\boldsymbol{x}}^q$. This means that $f(\widehat{\boldsymbol{x}}^1) = f(\widehat{\boldsymbol{x}}^2) = \cdots f(\widehat{\boldsymbol{x}}^q) = f^*$. Now let $\boldsymbol{x}$ be any convex combination of the above $\widehat{\boldsymbol{x}}^i$:

$$\boldsymbol{x} = \sum_{i=1}^{q} \alpha_i \widehat{\boldsymbol{x}}^i \quad \text{for } \alpha_i \geqq 0 \text{ and } \sum_i \alpha_i = 1$$

Then

$$\begin{aligned} f(\boldsymbol{x}^*) &= f(\alpha_1 \widehat{\boldsymbol{x}}^1 + \alpha_2 \widehat{\boldsymbol{x}}^2 + \cdots + \alpha_q \widehat{\boldsymbol{x}}^q) \\ &= \alpha_1 f(\widehat{\boldsymbol{x}}^1) + \alpha_2 f(\widehat{\boldsymbol{x}}^2) + \cdots + \alpha_q f(\widehat{\boldsymbol{x}}^q) = \sum_i \alpha_i f^* = f^* \end{aligned}$$

which establishes the result. ∎

Recall that a feasible solution is a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ with all $x_j \geqq 0$ such that

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \cdots + \boldsymbol{a}_n x_n = \boldsymbol{b}$$

Assume that we have found a set of k vectors that is linearly independent and that there exists a nonnegative combination of these vectors equal to $\boldsymbol{b}$. Let this set of (column) vectors be $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$. We then have the following theorem.

Theorem 2.1.4 If a set of $k \leqq m$ vectors, $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ can be found that is linearly independent such that

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \cdots + \boldsymbol{a}_k x_k = \boldsymbol{b}$$

and all $x_j \geqq 0$, then the point $\boldsymbol{x} = (x_1, x_2, \ldots, x_k, 0, \ldots, 0)$ is an extreme point of the convex set S of feasible solutions. Here $\boldsymbol{x}$ is an n-dimensional vector whose last $n - k$ components are zero.

Proof: If $\boldsymbol{x}$ is not an extreme point then it can be written as a convex combination of two other points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ in S. This means $\boldsymbol{x} = \alpha \boldsymbol{x}^1 + (1 - \alpha)\boldsymbol{x}^2$ for $0 < \alpha < 1$. Because all the components x_j of $\boldsymbol{x}$ are nonnegative and $\alpha \in (0, 1)$, the last $n - k$ components of $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ must also equal zero; that is,

$$\boldsymbol{x}^1 = \left(x_1^{(1)}, x_2^{(1)}, \ldots, x_k^{(1)}, 0, \ldots, 0\right) \text{ and } \boldsymbol{x}^2 = \left(x_1^{(2)}, x_2^{(2)}, \ldots, x_k^{(2)}, 0, \ldots, 0\right)$$

Given $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are feasible, we have $\boldsymbol{A}\boldsymbol{x}^1 = \boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{x}^2 = \boldsymbol{b}$, or more explicitly

$$\boldsymbol{a}_1 x_1^{(1)} + \boldsymbol{a}_2 x_2^{(1)} + \cdots + \boldsymbol{a}_k x_k^{(1)} = \boldsymbol{b} \text{ and } \boldsymbol{a}_1 x_1^{(2)} + \boldsymbol{a}_2 x_2^{(2)} + \cdots + \boldsymbol{a}_k x_k^{(2)} = \boldsymbol{b}$$

But $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is a linearly independent set and from linear algebra we know that $\boldsymbol{b}$ can be expressed as a unique linear combination of the elements $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$. This implies that $x_j = x_j^{(1)} = x_j^{(2)}$. Therefore, $\boldsymbol{x}$ cannot be expressed as a convex combination of two distinct points in S and so must be an extreme point. ∎

Theorem 2.1.5 If $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is an extreme point of S, then the vectors associated with positive x_j form a linearly independent set. It follows that at most m of the x_j are positive.

Proof: The proof is by contradiction. Let the first k components of $\boldsymbol{x}$ be nonzero so that $\sum_{j=1}^{k} \boldsymbol{a}_j x_j = \boldsymbol{b}$. Assume that $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is a linear dependent set. Then there exists a linear combination of these vectors that equal the zero vector,

$$\boldsymbol{a}_1 d_1 + \boldsymbol{a}_2 d_2 + \cdots + \boldsymbol{a}_k d_k = \boldsymbol{0} \tag{2.11}$$

with at least one $d_j \neq 0$. From the hypothesis of the theorem we have

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \cdots + \boldsymbol{a}_k x_k = \boldsymbol{b} \tag{2.12}$$

For some $\varepsilon > 0$, we multiply (2.11) by ε and add and subtract the result from (2.12) to obtain the two equations

$$\sum_{j=1}^{k} \boldsymbol{a}_j x_j + \varepsilon \sum_{j=1}^{k} \boldsymbol{a}_j d_j = \boldsymbol{b} \quad \text{and} \quad \sum_{j=1}^{k} \boldsymbol{a}_j x_j - \varepsilon \sum_{j=1}^{k} \boldsymbol{a}_j d_j = \boldsymbol{b}$$

This gives two solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ which may or may not satisfy the nonnegativity constraint $\boldsymbol{x} \geq \boldsymbol{0}$:

$$\boldsymbol{x}^1 = (x_1 + \varepsilon d_1, x_2 + \varepsilon d_2, \ldots, x_k + \varepsilon d_k) \text{ and } \boldsymbol{x}^2 = (x_1 - \varepsilon d_1, x_2 - \varepsilon d_2, \ldots, x_k - \varepsilon d_k)$$

But because $x_j > 0$, we can make ε as small as necessary to assure that the first k components of both $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are positive. So for an appropriate ε, $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are feasible solutions. But $\boldsymbol{x} = \frac{1}{2}\boldsymbol{x}^1 + \frac{1}{2}\boldsymbol{x}^2$ which contradicts the hypothesis that $\boldsymbol{x}$ is an extreme point. This implies that set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ cannot be linearly dependent.

Since every set of $m+1$ vectors in m-dimensional space is necessarily linearly dependent, we cannot have more than m positive x_j. If we did the proof of the main part of the theorem would imply that there exists $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m, \boldsymbol{a}_{m+1}$ linearly independent vectors. ∎

Without loss of generality it can be assumed that the technological matrix $\boldsymbol{A}$ associated with the linear programming problem always contains a set of m linearly independent vectors. If this property is not evident when a particular problem is being solved, the original set of vectors can be augmented by a set of m linearly independent vectors and the extended problem is solved instead. The details are explained in Section 2.4.1.

Corollary 2.1.1 Associated with every extreme point in S is a set of m linearly independent vectors from the given set $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$.

Proof: In Theorem 2.1.5 it was shown that there are $k \leqq m$ such vectors. For $k = m$, the corollary is proved. Assume that $k < m$ and that we can find only $\boldsymbol{a}_{k+1}$, $\boldsymbol{a}_{k+2}$, $\ldots, \boldsymbol{a}_r$ additional vectors such that the set $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_r\}$ for $r < m$ is linearly independent. This implies that the remaining $n - r$ vectors are dependent on $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$. But this contradicts the assumption that we always have a set of m linearly independent vectors in the given set $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. As a consequence, there must be m linearly independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ associated with every extreme point such that

$$\sum_{j=1}^{k} \boldsymbol{a}_j x_j + \sum_{j=k+1}^{m} \boldsymbol{a}_j 0 = \boldsymbol{b}$$

This completes the proof. ∎

Summarizing the results of this subsection, we have:

1. There is an extreme point of S at which the objective function takes on its minimum.
2. Every basic feasible solution corresponds to an extreme point of S; i.e., there is a unique mapping of basic feasible solutions to extreme points.
3. Every extreme point of S has m linearly independent vectors taken from the columns of $\boldsymbol{A}$ associated with it (the mapping of extreme points to bases may not be unique).

2.2 SIMPLEX METHOD

The simplex method for solving an LP in standard form generates a sequence of feasible points $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$ that terminates at an optimal solution under the assumption of nondegeneracy and boundedness. Because there exists an extreme point at which the solution occurs, the algorithm is designed so that each iterate $\boldsymbol{x}^{(k)}$ is an extreme point. Thus $n - m$ variables have zero value at $\boldsymbol{x}^{(k)}$ and are termed nonbasic variables. The remaining m variables take on nonnegative values (positive values under the nondegeneracy assumption) and are called basic variables. The simplex method makes systematic changes to these sets after each iteration to find the combination that gives the optimal solution. Each change is known as a pivot and is taken up next.

2.2.1 Pivoting

To obtain a firm grasp of the simplex procedure, it is essential that one first understands the process of pivoting in a set of simultaneous linear equations. Denoting the columns of $\boldsymbol{A}$ in (2.1b) by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, we get

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \cdots + \boldsymbol{a}_n x_n = \boldsymbol{b} \tag{2.13}$$

which expresses $\boldsymbol{b}$ as a linear combination of the $\boldsymbol{a}_j$ vectors. If $m < n$ and the vectors $\boldsymbol{a}_j$ span R^n then there is not a unique solution but a whole family of solutions. The vector $\boldsymbol{b}$ has a unique representation, however, as a linear combination of a given linearly independent subset of these vectors. The corresponding solution with $n - m$ x_j variables set to zero is a basic solution to (2.13).

Suppose now that we assume that the n variables are permutated so that the basic variables are the first m components of $\boldsymbol{x}$. Then we can write $\boldsymbol{x} = (\boldsymbol{x}_B, \boldsymbol{x}_N)$, where $\boldsymbol{x}_B$ and $\boldsymbol{x}_N$ refer to the basic and nonbasic variables, respectively. The matrix $\boldsymbol{A}$ can also be partitioned similarly into $\boldsymbol{A} = [\boldsymbol{B}, \boldsymbol{N}]$, where $\boldsymbol{B}$ is the $m \times m$ basis matrix and N is $m \times (n - m)$. The equations $\boldsymbol{Ax} = \boldsymbol{b}$ can thus be written

$$[\boldsymbol{B}, \boldsymbol{N}] \begin{pmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{pmatrix} = \boldsymbol{B}\boldsymbol{x}_B + \boldsymbol{N}\boldsymbol{x}_N = \boldsymbol{b} \tag{2.14}$$

Multiplying through by $\boldsymbol{B}^{-1}$ gives

$$\boldsymbol{x}_B + \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_N = \boldsymbol{B}^{-1}\boldsymbol{b} \tag{2.15}$$

which can be written in detached coefficient form giving rise to the following *canonical tableau*

$$\begin{array}{cccccccccc} 1 & 0 & \cdots & 0 & \alpha_{1,m+1} & \alpha_{1,m+2} & \cdots & \alpha_{1n} & \bar{b}_1 \\ 0 & 1 & \cdot & 0 & \alpha_{2,m+1} & \alpha_{2,m+2} & \cdot & \alpha_{2n} & \bar{b}_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 & \alpha_{m,m+1} & \alpha_{m,m+2} & \cdot & \alpha_{mn} & \bar{b}_m \end{array} \tag{2.16}$$

where the $m \times (n - m)$ matrix $(\boldsymbol{\alpha})$ is equivalently the matrix $\boldsymbol{B}^{-1}\boldsymbol{N}$ and the last column $\bar{\boldsymbol{b}}$ is $\boldsymbol{B}^{-1}\boldsymbol{b}$.

The first m vectors in (2.16) form a basis. Consequently, every other vector represented in the tableau can be expressed as a linear combination of these basis vectors by simply reading the coefficients down the corresponding column. For example,

$$\boldsymbol{a}_j = \alpha_{1j}\boldsymbol{a}_1 + \alpha_{2j}\boldsymbol{a}_2 + \cdots + \alpha_{mj}\boldsymbol{a}_m \tag{2.17}$$

The tableau can be interpreted as giving the representations of the vectors $\boldsymbol{a}_j$ in terms of the basis; the jth column of the tableau is the representation for the vector $\boldsymbol{a}_j$. Moreover, the expression for $\boldsymbol{b}$ in terms of the basis is given in the last column.

Now consider the operation of replacing one member of the basis by an outside column. Suppose for example that we wish to replace the basis vector $\boldsymbol{a}_p$, $1 \leq p \leq m$, by the vector $\boldsymbol{a}_q$. Provided that the first m vectors with $\boldsymbol{a}_p$ replaced by $\boldsymbol{a}_q$ are linearly independent these vectors constitute a basis and every vector can be expressed as a linear combination of this new basis. To find the new representations of the vectors we must update the tableau. The linear independence condition holds if and only if $\alpha_{pq} \neq 0$.

Any vector $\boldsymbol{a}_j$ can be written in terms of the old array using (2.17); for $\boldsymbol{a}_q$ we have

$$\boldsymbol{a}_q = \sum_{\substack{i=1 \\ i \neq p}}^{m} \alpha_{iq} \boldsymbol{a}_i + \alpha_{pq} \boldsymbol{a}_p$$

Solving for $\boldsymbol{a}_p$,

$$\boldsymbol{a}_p = \frac{1}{\alpha_{pq}} \boldsymbol{a}_q - \sum_{\substack{i=1 \\ i \neq p}}^{m} \frac{\alpha_{iq}}{\alpha_{pq}} \boldsymbol{a}_i \tag{2.18}$$

and substituting (2.18) into (2.17) gives

$$\boldsymbol{a}_j = \sum_{\substack{i=1 \\ i \neq p}}^{m} \left(\alpha_{ij} - \frac{\alpha_{iq}}{\alpha_{pq}} \alpha_{pj} \right) \boldsymbol{a}_i + \frac{\alpha_{pj}}{\alpha_{pq}} \boldsymbol{a}_q \tag{2.19}$$

Denoting the coefficients of the new tableau by α'_{ij}, we obtain immediately from (2.19)

$$\begin{cases} \alpha'_{ij} = \alpha_{ij} - \dfrac{\alpha_{iq}}{\alpha_{pq}} \alpha_{pj}, & i \neq p \\ \alpha'_{pj} = \dfrac{\alpha_{pj}}{\alpha_{pq}} \end{cases}$$

In linear algebra terms, the pivoting operation described above is known as *Gauss-Jordan elimination.*

If a system of equations is not originally given in canonical form, we can put it into this form by adjoining the m unit vectors to the tableau and, starting with these vectors as the basis, successively replace each of them with columns of $\boldsymbol{A}$ using the pivot operation.

Example 2.2.1 Suppose that we wish to solve the system of simultaneous equations

$$\begin{aligned} x_1 + x_2 - x_3 &= 5 \\ 2x_1 - 3x_2 + x_3 &= 3 \\ -x_1 + 2x_2 - x_3 &= -1 \end{aligned}$$

To obtain an initial basis, we add three variables x_4, x_5 and x_6 to get

$$\begin{aligned} x_1 + x_2 - x_3 + x_4 &= 5 \\ 2x_1 - 3x_2 + x_3 + x_5 &= 3 \\ -x_1 + 2x_2 - x_3 + x_6 &= -1 \end{aligned}$$

Now to find the basic solution in terms of the variables x_1, x_2 and x_3, we set up the tableau below:

x_1	x_2	x_3	x_4	x_5	x_6	$\bar{\boldsymbol{b}}$
(1)	1	−1	1	0	0	5
2	−3	1	0	1	0	3
−1	2	−1	0	0	1	−1

The first pivot element, which is circled, is $\alpha_{11} = 1$ and corresponds to the replacement of x_4 by x_1 as a basic variable. After pivoting we obtain the array

x_1	x_2	x_3	x_4	x_5	x_6	$\bar{\boldsymbol{b}}$
1	1	−1	1	0	0	5
0	(−5)	3	−2	1	0	−7
0	3	−2	1	0	1	4

where the next pivot element $\alpha_{22} = -5$ indicates that x_5 is to be replaced by x_2. Continuing through two additional tableaus gives us the final basic solution $x_1 = 4$, $x_2 = 2$, $x_3 = 1$ with nonbasic variables $x_4 = x_5 = x_6 = 0$.

2.2.2 Determining the Leaving Variable

An arbitrary selection of m linearly independent columns from $\boldsymbol{A}$ does not guarantee that the corresponding basic variables will satisfy the nonnegativity condition $x \geqq \mathbf{0}$. Even if we start with a known feasible basis, the pivot operation, which takes one basis into another will not in general preserve feasibility. Special conditions must be met to maintain feasibility from one iteration to the next. How this is done lies at the heart of the simplex method.

Suppose we have the basic feasible solution (BFS) $\boldsymbol{x} = (x_1, x_2, \ldots, x_m, 0, \ldots, 0)$ or the equivalent representation

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_m\boldsymbol{a}_m = \boldsymbol{b} \tag{2.20}$$

Under the nondegeneracy assumption, $x_i > 0$, $i = 1, \ldots, m$. Now let us represent the vector $\boldsymbol{a}_q$, $q > m$, in terms of the current basis as follows:

$$\boldsymbol{a}_q = \alpha_{1q}\boldsymbol{a}_1 + \alpha_{2q}\boldsymbol{a}_2 + \cdots + \alpha_{mq}\boldsymbol{a}_m \tag{2.21}$$

where recall that $\boldsymbol{\alpha}_q = (\alpha_{1q}, \ldots, \alpha_{mq})^T$ is the $(q - m + 1)$st column of $\boldsymbol{B}^{-1}\boldsymbol{N}$. Multiplying (2.21) by the parameter $\theta \geqq 0$ and subtracting it from (2.20) gives

$$(x_1 - \theta\alpha_{1q})\boldsymbol{a}_1 + (x_2 - \theta\alpha_{2q})\boldsymbol{a}_2 + \cdots + (x_m - \theta\alpha_{mq})\boldsymbol{a}_m + \theta\boldsymbol{a}_q = \boldsymbol{b} \tag{2.22}$$

For $\theta \geqq 0$ we have $\boldsymbol{b}$ as a linear combination of at most $m + 1$ vectors. (In fact, (2.22) is a line in m-space which, we will see in Section 2.3.2, represents an edge of

the polytope S.) For $\theta = 0$ we have the original BFS; as θ is increased from zero the coefficient of $\boldsymbol{a}_q$ increases, and for small enough θ, (2.22) gives a feasible but nonbasic solution. The coefficients of the other vectors will either increase or decrease linearly as θ increases. If any decrease, to maintain feasibility we may set θ equal to the value corresponding to the first instance where one (or more) of the coefficients vanishes. That is

$$\theta = \min_i \left\{ \frac{x_i}{\alpha_{iq}} \;\middle|\; \alpha_{iq} > 0 \right\} \tag{2.23}$$

Placing the value of θ found in (2.23) in (2.22) yields a new BFS with the vector $\boldsymbol{a}_q$ replaced by the vector $\boldsymbol{a}_p$, where p corresponds to the minimizing index in (2.23). The calculation of θ is known as the *minimum ratio test*. If the minimum is achieved at more than a single index i, then the new solution is degenerate and any of the vectors with zero coefficient can be regarded as the one that left the basis.

If none of the α_{iq}'s are positive, then all coefficients in (2.22) increase (or remain the same) as θ is increased, and no new basic feasible solution emerges. This means that there are feasible solutions to (2.1b)–(2.1c) that have arbitrarily large coefficients so that the set S is unbounded. As we shall see, this case is of special significance in the simplex method.

2.2.3 Moving toward Optimality

To find an optimal solution we begin with any basic feasible solution and then attempt to find another with a smaller objective function value f. Suppose that we have a BFS given by (2.15) which is equivalent to

$$\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}, \quad \boldsymbol{x}_N = \boldsymbol{0} \tag{2.24}$$

Let $\boldsymbol{c} = (\boldsymbol{c}_B, \boldsymbol{c}_N)$, and thus

$$f = \boldsymbol{c}_B\boldsymbol{x}_B + \boldsymbol{c}_N\boldsymbol{x}_N \tag{2.25}$$

Using (2.15) to eliminate $\boldsymbol{x}_B$ from (2.25) gives

$$f = \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b} + (\boldsymbol{c}_N - \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{N})\boldsymbol{x}_N \tag{2.26}$$

Substituting $\boldsymbol{x}_N = \boldsymbol{0}$ in (2.26) yields $f = \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b}$ as the value of the basic solution given by (2.24). For convenience, let $f^0 \triangleq \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b}$, $\boldsymbol{x}_B \triangleq (x_{B_1}, \ldots, x_{B_m})$, and Q be the index set of columns in $\boldsymbol{N}$. Then (2.15) and (2.26) can be written as

$$x_{B_i} = \bar{b}_i - \sum_{j \in Q} \alpha_{ij} x_j \tag{2.27a}$$

$$f = f^0 + \sum_{j \in Q} \bar{c}_j x_j \tag{2.27b}$$

where $\bar{c}_j = c_j - \boldsymbol{c}_B \boldsymbol{B}^{-1} \boldsymbol{a}_j$, $j \in Q$; that is, the coefficients of the nonbasic variables in (2.26). These values are known as the *reduced cost coefficients* or the *relative cost coefficients*. The solution given by (2.24) is obtained by setting $x_j = 0$ for all $j \in Q$ in (2.27).

Assume that $\boldsymbol{x}_B$ is nondegenerate and $\bar{c}_j < 0$ for some $j \in Q$, say $j = q$. Then by increasing x_q and keeping all other nonbasic variables fixed at zero, f decreases linearly with slope $\bar{c}_q$. Also, x_{B_i} is a linear function of x_q with slope $-\alpha_{iq}$. As was shown in Section 2.2.2, if $\alpha_{iq} > 0$, then $x_{B_i} \geqq 0$ as long as $x_q \leqq \bar{b}_i / \alpha_{iq} \triangleq \theta_{iq}$. When $x_q = \theta_{iq}$, $x_{B_i} = 0$.

Let x_{B_p} be any basic variable with

$$0 < \theta_{pq} = \min_{i=1,\ldots,m} \{\theta_{iq} : \alpha_{iq} > 0\}$$

Then by keeping all nonbasic variables equal to zero except x_q, letting $\theta \triangleq \theta_{pq}$, and setting

$$\begin{aligned} x_q &= \theta \\ x_{B_i} &= \bar{b}_i - \alpha_{iq}\theta, \quad i = 1, \ldots, m \end{aligned}$$

a new BFS is obtained with $x_q > 0$, $x_{B_p} = 0$, and $f = f^0 + \bar{c}_q \theta$. Because $\theta > 0$ and $\bar{c}_q < 0$, f is decreased. Observe that if there is no q such that $\alpha_{iq} > 0$, then x_q can be made arbitrarily large while the basic variables remain nonnegative. From (2.27b) we see that the objective function f can thus be made arbitrarily small implying that the problem is unbounded.

Finding the new BFS corresponds to solving the pth equation of (2.27a) for x_q, eliminating x_q from the other equations for $i \neq p$, and then setting $x_{B_p} = 0$ and $x_j = 0$ for all $j \in Q \setminus \{q\}$. Thus whenever there is a nondegenerate BFS with $\bar{c}_j < 0$ for some $j \in Q$, say $j = q$, and $\alpha_{iq} > 0$ for at least one i, a better BFS can be found by exchanging one column of $\boldsymbol{N}$ for one of $\boldsymbol{B}$. This result is summarized in the following theorem.

Theorem 2.2.1 Given a nondegenerate BFS with corresponding objective function value f^0, suppose that for some j there holds $\bar{c}_j < 0$ for $j \in Q$. Then there is a feasible solution with objective function value $f < f^0$. If the column $\boldsymbol{a}_j$ can be substituted for some vector in the original basis to yield a new BFS, this new solution will have $f < f^0$. If no $\boldsymbol{a}_j$ can found, then the solution set S is unbounded and the objective function can be made arbitrarily small (toward minus infinity).

On the other hand, suppose that $\boldsymbol{B}$ is a basis matrix which yields vector $\boldsymbol{x}^*$ with $\bar{c}_j \geqq 0$ for all $j \in Q$ and corresponding f^*. From (2.27b) the objective function can be written as

$$f^* = \boldsymbol{c}_B \boldsymbol{B}^{-1} \boldsymbol{b} + \sum_{j \in Q} \bar{c}_j x_j^* \tag{2.28}$$

which does not contain $\boldsymbol{x}_B$. From the fact that $\bar{c}_j \geqq 0$ and $x_j^* \geqq 0$ for all $j \in Q$, we can conclude that $\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b}$ is a lower bound on f. Because $\boldsymbol{x}_B^* = \boldsymbol{B}^{-1}\boldsymbol{b}$, $\boldsymbol{x}_N^* = \boldsymbol{0}$ is a feasible and achieves this lower bound, it is optimal. Thus we have proved

Theorem 2.2.2 The basic solution given in (2.27a) is optimal if

i) $x_{B_i} \geqq 0$, $i = 1, \ldots, m$ (feasibility)

ii) $\bar{c}_j \geqq 0$, for all $j \in Q$.

Finding improved basic feasible solutions, as described above, is essentially the simplex algorithm. Extensive experience has shown that starting with an initial basis, an optimal solution can be found in about m, or perhaps $3m/2$ pivot operations. As such, if m is much smaller than n, that is, the matrix $\boldsymbol{A}$ has far fewer rows than columns, pivots will occur in only a small fraction of the columns during the course of optimization. To take advantage of this fact, the revised form of the simplex method was developed and will be described below. It starts with the inverse $\boldsymbol{B}^{-1}$ of the current basis, and the current solution $\boldsymbol{x}_B = \bar{\boldsymbol{b}} = \boldsymbol{B}^{-1}\boldsymbol{b}$.

Algorithm

Step 1 (Optimality test) Calculate the current reduced cost coefficients $\bar{\boldsymbol{c}} = \boldsymbol{c}_N - \boldsymbol{c}_B\boldsymbol{B}^{-1}N$. This can be done most efficiently by first calculating the vector $\boldsymbol{\lambda}^T = \boldsymbol{c}_B\boldsymbol{B}^{-1}$ and then the reduced cost vector $\bar{\boldsymbol{c}} = \boldsymbol{c}_N - \boldsymbol{\lambda}^T\boldsymbol{N}$. If $\bar{\boldsymbol{c}} \geqq \boldsymbol{0}$ stop, the current solution is optimal.

Step 2 (Entering variable) Determine which vector $\boldsymbol{a}_q$ is to enter the basis by selecting the most negative cost coefficient; calculate $\boldsymbol{\alpha}_q = \boldsymbol{B}^{-1}\boldsymbol{a}_q$ which gives the vector $\boldsymbol{a}_q$ in terms of the current basis.

Step 3 (Leaving variable) If no $\alpha_{iq} > 0$, stop; the problem is unbounded. Otherwise, calculate the ratios $\bar{b}_i/\alpha_{iq}$ for $\alpha_{iq} > 0$ to determine which vector is to leave the basis. Break ties arbitrarily.

Step 4 (Pivoting) Update $\boldsymbol{B}^{-1}$ and the current solution and $\boldsymbol{B}^{-1}\boldsymbol{b}$. Return to Step 1.

The computations in Step 1 where the new set of reduced cost coefficients is found are known as the *pricing out* operation. In Step 4, the simplest way to update $\boldsymbol{B}^{-1}$ is by applying the usual pivot operations to an array consisting of $\boldsymbol{B}^{-1}$ and $\boldsymbol{\alpha}_q$, where the pivot element is the appropriate component of $\boldsymbol{\alpha}_q$. Of course $\boldsymbol{B}^{-1}\boldsymbol{b}$ can be updated at the same time by adjoining it as another column.

To begin the procedure, one needs an initial basic feasible solution and the inverse of the accompanying basis. In most problems, the initial basis is taken to be the identity matrix of appropriate dimension resulting from the adjoining of slack, surplus or artificial variables to the original formulation (artificial variables are discussed in Section 2.4.1). In practice, more elaborate procedures are used in Step 2 to determine

the entering variable and to take into account the possibility of degeneracy, and in Step 4 to keep track of $\boldsymbol{B}^{-1}$ (see [B25] and [S17], respectively).

Example 2.2.2 To illustrate the computations of the revised simplex method we will solve the following problem

$$\begin{array}{rr} \min & -3x_1 - x_2 - 3x_3 \\ \text{subject to} & 2x_1 + x_2 + x_3 \leqq 2 \\ & x_1 + 2x_2 + 3x_3 \leqq 5 \\ & 2x_1 + 2x_2 + x_3 \leqq 6 \\ & x_1, x_2, x_3 \geqq 0 \end{array}$$

To put the constraints into standard form, we introduce three slack variables x_4, x_5 and x_6 which leads to the following tableau

x_1	x_2	x_3	x_4	x_5	x_6	$\boldsymbol{b}$
2	1	1	1	0	0	2
1	2	3	0	1	0	5
2	2	1	0	0	1	6

with objective function coefficients $\boldsymbol{c} = (-3, -1, -3, 0, 0, 0)$. We start with an initial BFS and corresponding $\boldsymbol{B}^{-1}$ as shown in the inverse tableau below

Basic variable	$\boldsymbol{B}^{-1}$			$\boldsymbol{x}_B$
4	1	0	0	2
5	0	1	0	5
6	0	0	1	6

Beginning at Step 1, we compute $\boldsymbol{\lambda}^T = (0,0,0)\boldsymbol{B}^{-1} = (0,0,0)$ and then $\bar{\boldsymbol{c}} = \boldsymbol{c}_N - \boldsymbol{\lambda}^T\boldsymbol{N} = (-3,-1,-3)$. At Step 2, rather than picking the most negative reduced cost coefficient, we decide to bring $\boldsymbol{a}_2$ into the basis (this simplifies the calculations and demonstrates that any $\bar{c}_j < 0$ is acceptable). Its current representation is found by multiplying by $\boldsymbol{B}^{-1} = I_3$, giving

Basic variable	$\boldsymbol{B}^{-1}$			$\boldsymbol{x}_B$	$\boldsymbol{\alpha}_2$
4	1	0	0	2	(1)
5	0	1	0	5	2
6	0	0	1	6	2

After performing the minimum ratio test in the usual manner, we select the pivot element as indicated. The updated inverse tableau becomes

Basic variable	$\boldsymbol{B}^{-1}$			$\boldsymbol{x}_B$	$\boldsymbol{\alpha}_3$
2	1	0	0	2	1
5	−2	1	0	1	(1)
6	−2	0	1	2	−1

then $\boldsymbol{\lambda}^T = (-1,0,0)\boldsymbol{B}^{-1} = (-1,0,0)$ and $\bar{c}_1 = -1$, $\bar{c}_3 = -2$, $\bar{c}_4 = 1$. We select $\boldsymbol{a}_3$ to enter. The updated $\boldsymbol{\alpha}_3$ is given in the above tableau. Pivoting as indicated, we obtain

Basic variable	$\boldsymbol{B}^{-1}$			$\boldsymbol{x}_B$	$\boldsymbol{\alpha}_1$
2	1	−1	0	1	(5)
3	−2	1	0	1	−3
6	−4	1	1	3	−5

which leads to $\boldsymbol{\lambda}^T = (-1,-3,0)\boldsymbol{B}^{-1} = (3,-2,0)$ and $\bar{c}_1 = -7$, $\bar{c}_4 = -3$, $\bar{c}_5 = 2$. We select $\boldsymbol{a}_1$ to enter and pivot on the indicated element, giving

Basic Variable	$\boldsymbol{B}^{-1}$			$\boldsymbol{x}_B$
1	3/5	−1/5	0	1/5
3	−1/5	2/5	0	8/5
6	−1	0	1	4

with $\boldsymbol{\lambda}^T = (-3,-3,0)\boldsymbol{B}^{-1} = (-6/5,-3/5,0)$ and $\bar{c}_2 = 7/5$, $\bar{c}_4 = 6/5$, $\bar{c}_5 = 3/5$. Because all the $\bar{c}_j$ are nonnegative, we conclude that the solution $\boldsymbol{x}^* = (1/5,0,8/5,0,0,4)$ is optimal.

2.2.4 Degeneracy and Cycling

A degenerate basic feasible solution occurs with, for some p and q, $\bar{c}_q < 0$, $\bar{b}_p = 0$, and $\alpha_{pq} > 0$. By introducing x_q into the basis, x_{B_p} (or some other basic variable equal to 0) must leave the basis and although a new basis is obtained, the solution and objective function value remain the same. The new basic solution $x_{B_i} = \bar{b}_i$, $i \neq p$, $x_q = 0$, $x_j = 0$ otherwise, is also degenerate. Examples have been constructed in which a finite sequence of degenerate bases obtained by the simplex algorithm produces a cycle where no progress is made toward the solution; that is, the sequence $\boldsymbol{B}^1, \ldots, \boldsymbol{B}^{k-1}, \boldsymbol{B}^k$ occurs with $\boldsymbol{B}^k = \boldsymbol{B}^1$.

This unwelcome phenomenon is called *cycling*. A wealth of empirical evidence indicates that cycling does not occur with any regularity in real applications so it mostly

of academic interest for linear programming. Nevertheless, the simplex algorithm can be modified so that it will never occur. The procedure used is of primary importance in some integer programming algorithms where LPs are repeatedly solved as subproblems, and involves specializing Step 3 of the revised simplex method. Recall that in Step 3 a variable is chosen to leave the basis. When there is more than one candidate, an arbitrary choice is made. By giving a special rule for breaking ties, the possibility of cycling can be avoided. Before presenting the rule, some new definitions are introduced.

A vector $\boldsymbol{v} \neq \boldsymbol{0}$ is called *lexicographically positive* if its first nonzero component is positive. The notation $\boldsymbol{v} \succ \boldsymbol{0}$ is used to denote this situation. For example $(0, 0, 3, -9, 4) \succ \boldsymbol{0}$. A vector $\boldsymbol{v}$ is said to be *lexicographically greater* than a vector $\boldsymbol{u}$ if $\boldsymbol{v} - \boldsymbol{u} \succ \boldsymbol{0}$. A sequence of vectors $\{\boldsymbol{v}^t\}$ is called *lexicographically increasing* if $\boldsymbol{v}^{t+1} - \boldsymbol{v}^t \succ \boldsymbol{0}$ for all t. If $-\boldsymbol{v} \succ \boldsymbol{0}$, $\boldsymbol{v}$ is called *lexicographically negative* (denoted by $\boldsymbol{v} \prec \boldsymbol{0}$) and if $\boldsymbol{v} - \boldsymbol{u} \prec \boldsymbol{0}$, $\boldsymbol{v}$ is *lexicographically smaller* than $\boldsymbol{u}$. The notation $\boldsymbol{v} \succeq \boldsymbol{0}$ means that $\boldsymbol{v} = \boldsymbol{0}$ or $\boldsymbol{v} \succ \boldsymbol{0}$.

The changes required in the simplex algorithm to avoid cycling can best be described using the tableau below which represents an initial BFS.

								x_j	
x_{B_1}	$\bar{b}_1$	1	$\cdots$	0	$\cdots$	0	$\cdots$	α_{1j}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	
x_{B_i}	$\bar{b}_i$	0	$\cdots$	1	$\cdots$	0	$\cdots$	α_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$	
x_{B_m}	$\bar{b}_m$	0	$\cdots$	0	$\cdots$	1	$\cdots$	α_{mj}	$\cdots$
$-f$	f^0	0	$\cdots$	0	$\cdots$	0	$\cdots$	$\bar{c}_j$	$\cdots$

In the tableau, an mth order identity matrix is inserted between the solution column and the columns corresponding to the nonbasic variables. The last row contains the objective function. The m unit columns are not used to determine the variable to enter the basis; however, at each iteration, they are transformed by applying the standard pivoting rules given in the simplex method to update $\boldsymbol{B}^{-1}$. As will be shown, the only purpose of these columns is to identify the variable to leave the basis when the choice at Step 3 in not unique.

Let $\boldsymbol{v}^0 = (f^0, 0, \ldots, 0)$ and $\boldsymbol{v}^i = (\bar{b}_i, 0, \ldots, 1, \ldots, 0)$, $i = 1, \ldots, m$, be the coefficients in the first $m + 1$ columns of the ith row of the above tableau. Because $\bar{b}_i \geqq 0$, it follows that $\boldsymbol{v}^i \succ \boldsymbol{0}$, $i = 1, \ldots, m$.

Suppose that $\bar{c}_q < 0$ and x_q is chosen to enter the basis. Let $E(q) = \{i : 1 \leqq i \leqq m, \alpha_{iq} > 0\}$ and $\boldsymbol{u}^i = \boldsymbol{v}^i/\alpha_{iq}$ for all $i \in E(q)$. Assume that $\boldsymbol{u}^p$ is the lexicographically

smallest of these vectors, that is,

$$\boldsymbol{u}^p = \operatorname*{lexmin}_{i \in E(q)} \ \boldsymbol{u}^i \tag{2.29}$$

The vector $\boldsymbol{u}^p$ must be unique because the $\boldsymbol{v}^i$'s are linearly independent. We now choice x_{B_p} to leave the basis. Note that x_{B_p} determined from (2.29) satisfies the minimum ratio rule specified in the simplex algorithm since

$$\boldsymbol{u}^p = \operatorname*{lexmin}_{i \in E(q)} \ \boldsymbol{u}^i \quad \Longrightarrow \quad \frac{\bar{b}_p}{\alpha_{pq}} = \min_{i \in E(q)} \ \frac{\bar{b}_i}{\alpha_{iq}}$$

Pivoting on row p and column q yields

$$\begin{aligned} \widehat{\boldsymbol{v}}^p &= \frac{\boldsymbol{v}^p}{\alpha_{pq}} \\ \widehat{\boldsymbol{v}}^i &= \boldsymbol{v}^i - \frac{\alpha_{iq}}{\alpha_{pq}} \boldsymbol{v}^p = \boldsymbol{v}^i - \alpha_{iq} \widehat{\boldsymbol{v}}^p, \quad i \neq p \end{aligned}$$

Since $\boldsymbol{v}^p \succ \boldsymbol{0}$ and $\alpha_{pq} > 0$, $\widehat{\boldsymbol{v}}^p \succ \boldsymbol{0}$. If $\alpha_{iq} \leq 0$, then $-\alpha_{iq} \widehat{\boldsymbol{v}}^p \succeq \boldsymbol{0}$ and $\widehat{\boldsymbol{v}}^i \succ \boldsymbol{0}$. If $\alpha_{iq} > 0$, we have

$$\boldsymbol{u}^i - \boldsymbol{u}^p = \frac{\boldsymbol{v}^i}{\alpha_{iq}} - \widehat{\boldsymbol{v}}^p \succ \boldsymbol{0} \tag{2.30}$$

from (2.29). Multiplying (2.30) by $\alpha_{iq} > 0$ yields $\widehat{\boldsymbol{v}}^i \succ \boldsymbol{0}$.

Thus $\widehat{\boldsymbol{v}}^i$, $i = 1, \ldots, m$, are lexicographically positive as well as linearly independent since adding multiples of one vector to another does not destroy independence. By induction, under the pivot row selection rule given by (2.29), $\widehat{\boldsymbol{v}}^i$, $i = 1, \ldots, m$, remains linearly independent and lexicographically positive at each iteration.

At the kth iteration the tableau is completely determined by the basis $\boldsymbol{B}^k$. Thus if tableaus k and t are such that $\boldsymbol{v}_k^0 \neq \boldsymbol{v}_t^0$, then $\boldsymbol{B}^k \neq \boldsymbol{B}^t$. Note that if the sequence $\{\boldsymbol{v}_t^0\}$ is lexicographically increasing, no distinct t_1 and t_2 exist such that $\boldsymbol{v}_{t_1}^0 = \boldsymbol{v}_{t_2}^0$. However,

$$\boldsymbol{v}_{t+1}^0 = \boldsymbol{v}_t^0 - \frac{\bar{c}_q^t}{\alpha_{pq}^t} \boldsymbol{v}_t^p = \boldsymbol{v}_t^0 - \bar{c}_q^t \boldsymbol{v}_{t+1}^p$$

so that $\boldsymbol{v}_{t+1}^p \succ \boldsymbol{0}$ and $\bar{c}_q^t < 0$ imply $\boldsymbol{v}_{t+1}^0 \succ \boldsymbol{v}_t^0$.

Example 2.2.3 Consider the following tableau which indicates that x_5 is the entering variable.

		$\bar{\boldsymbol{b}}$				x_4	x_5
							$\downarrow$
	x_1	0	1	0	0	1	1
	x_2	0	0	1	0	2	4
$\leftarrow$	x_3	0	0	0	1	-1	3
	$-f$	0	0	0	0	2	-1

The ratio test at Step 3 of the simplex algorithm would permit the removal of either x_1, x_2 or x_3 because $\theta_{i5} = \bar{b}_i/\alpha_{i5} = 0$ for $i = 1, 2, 3$. Nevertheless, $\boldsymbol{u}^3 = (0, 0, 0, 1/3) \prec \boldsymbol{u}^2 = (0, 0, 1/4, 0) \prec \boldsymbol{u}^1 = (0, 1, 0, 0)$ so x_3 is chosen as the departing variable. Observe that in the updated tableau given below, $\widehat{\boldsymbol{v}}^i \succ \mathbf{0}$ and $\widehat{\boldsymbol{v}}^0 - \boldsymbol{v}^0 \succ \mathbf{0}$. The reader can verify that if x_1 or x_2 were chosen to leave, then $\widehat{\boldsymbol{v}}^3 \prec \mathbf{0}$.

	$\bar{\boldsymbol{b}}$				x_4	x_3
x_1	0	1	0	$-1/3$	$4/3$	$-1/3$
x_2	0	0	1	$-4/3$	$10/3$	$-4/3$
x_5	0	0	0	$1/3$	$-1/3$	$1/3$
$-f$	0	0	0	$1/3$	$5/3$	$1/3$

In addition to the lexicographic rule for preventing cycling, a number of other techniques have been proposed. These are discussed, for example, by Murty [M18] but are likewise computational inefficient. As a practical matter, cycling occurs only in rare instances so the designers of most commercial LP codes have simply ignored it. A second reason for ignoring cycling relates to computer round-off errors. In this regard, it can be argued that the updated right-hand-side values $\bar{b}_i$ are usually perturbed from their actual values due to round-off errors so one rarely encounters exact zero values for $\bar{b}_i$. In fact, the lexicographic rule has an equivalent interpretation as a perturbation technique in which the right-hand-side values are modified slightly to make the polytope nondegenerate.

The only real case of practical concern where cycling may present difficulties is in network optimization. For those linear programs that have a network structure, specialized codes are available that provide significant computational advantage over the simplex method. Fortunately, cycling prevention rules are easy to implement in these codes.

The final issue to address concerns a phenomenon related to cycling where the simplex algorithm goes through an exceedingly long (though finite) sequence of degenerate pivots whose number may be exponential in the size (m and n) of the problem. This phenomenon is known as *stalling*. The term arises because with increasing problem size, the algorithm can spend an enormous amount of time at a degenerate vertex before verifying optimality or moving on. Besides preventing cycling one would like

to preclude stalling by guaranteeing that the length of a sequence of degenerate pivots is bounded from above by a polynomial in m and n. This issue is further discussed in [B15].

2.3 GEOMETRY OF SIMPLEX METHOD

There are two ways, in general, to interpret the simplex method geometrically. The first, and perhaps the most natural, is in *activity space* where $\boldsymbol{x}$ is represented. As discussed in Section 2.1.3, the feasible region is viewed directly as a convex set, and basic feasible solutions are extreme points. Adjacent extreme points are points that lie on a common edge.

The second geometrical interpretation is in *requirements space*, the space where the columns of $\boldsymbol{A}$ and $\boldsymbol{b}$ are represented. Equation (2.13) depicts the fundamental relation. An example for $m = 2$ and $n = 4$ is shown in Fig. 2.2. Here a feasible solution is defined by $\boldsymbol{b}$ lying is a 'pos cone' associated with the columns of $\boldsymbol{A}$. As described in [M18], a nondegenerate basic feasible solution will use exactly m positive weights. In the figure a basic feasible solution can be constructed with positive weights on $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ because $\boldsymbol{b}$ lies between them. Such a solution cannot be constructed using $\boldsymbol{a}_1$ and $\boldsymbol{a}_4$. Suppose we start with $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ as the initial basis. Then an adjacent basis is found by bringing in some other vector. If $\boldsymbol{a}_3$ is brought in, then clearly $\boldsymbol{a}_2$ must leave. On the other hand, if $\boldsymbol{a}_4$ is introduced, $\boldsymbol{a}_1$ must leave.

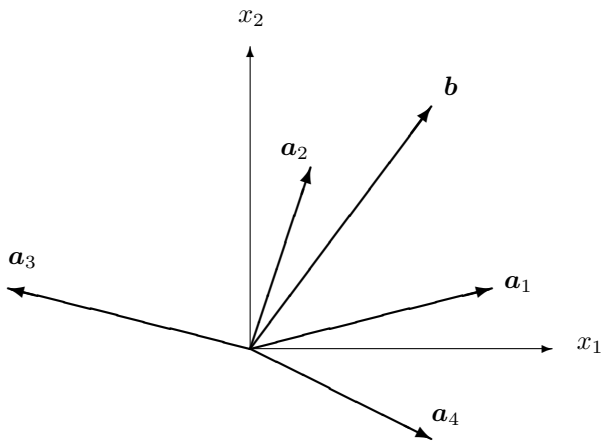

Figure 2.2 Constraint representation in requirements space

2.3.1 Finiteness of Algorithm

Consider again the LP in standard form (2.1) where $\boldsymbol{A}$ is an $m \times n$ matrix of rank m. If the problem has an optimal feasible solution, then it has an optimal BFS by Theorem 2.1.1. Therefore, it is sufficient to search among the finite number of BFSs of (2.1) for an optimum. This is what the simplex method does.

If (2.1) is a nondegenerate LP, every BFS has exactly m positive components and hence, has a unique associated basis. In the case, when the simplex algorithm is used to solve (2.1), the minimum ratio will turn out to be strictly positive in every pivot step so the objective function will decrease at each iteration. This means that a basis that appears once in the course of the algorithm can never reappear. Because the total number of bases is finite, the algorithm must terminate after a finite number of pivot steps with either an optimal basic vector or by satisfying the unboundedness criterion .

Now suppose that (2.1) is a degenerate LP so that when applying the simplex method, a degenerate pivot may occur. The next pivot step in the algorithm might also turn out to be degenerate leading to cycling. During these pivot steps, the algorithm is just moving among bases, all of which are associated with the same extreme point. Nevertheless, with a pivot row choice rule for resolving degeneracy, such as the lexicographic minimum ratio rule (2.29), it is possible to avoid returning to a previously encountered basis. This leads to the following theorem.

Theorem 2.3.1 Starting with a primal feasible basis, the simplex algorithm (with some technique for resolving degeneracy) finds, after a finite number of pivots, one of the following:

1. An optimal feasible basis whose corresponding canonical tableau satisfies the optimality criterion.
2. A primal feasible basis whose canonical tableau satisfies the unboundedness criterion.

2.3.2 Adjacency and Bounded Edges

In this subsection we see how the simplex algorithm walks from one feasible vertex in S to another before a termination criterion is satisfied. Consider the convex polytope depicted in Fig. 2.3. Evidently every point on the line segment joining the extreme points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ cannot be expressed as a convex combination of any pair of points in the polytope that are not on this line segment. This, however, is not true of points on the line segment joining the extreme points $\boldsymbol{x}^1$ and $\boldsymbol{x}^3$. Extreme points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are known as adjacent extreme points of the convex polytope. The extreme points $\boldsymbol{x}^1$ and $\boldsymbol{x}^3$ are not adjacent.

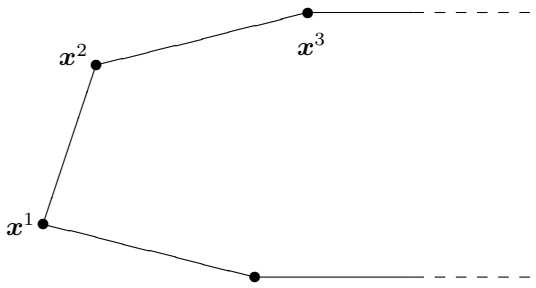

Figure 2.3 Adjacent edges of convex polytope

Definition 2.3.1 (*Geometric characterization of adjacency*) Two extreme points $\tilde{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$ of a convex polytope S are said to be adjacent iff every point $\bar{\boldsymbol{x}}$ on the line segment joining them has the property that if $\bar{\boldsymbol{x}} = \alpha \boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2$, where $0 < \alpha < 1$ and $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, then both $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ must themselves be on the line segment joining $\tilde{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$. The line segment joining a pair of adjacent extreme points is called a *bounded edge* or *edge* of the convex polytope.

Definition 2.3.2 (*Algebraic characterization of adjacency*) When dealing with the convex polytope S associated with an LP, there is a simple characterization of adjacency in terms of the rank of a set of vectors. That is, two distinct extreme points of S, $\tilde{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$, are adjacent iff the rank of the set $\{\boldsymbol{a}_j : j$ such that either $\tilde{x}_j$ or $\widehat{x}_j$ or both are $> 0\}$ is one less than its cardinality. In other words, this set is linearly dependent and contains a column vector such that when this column vector is deleted from the set, the remaining set is linearly independent.

Theorem 2.3.2 Let S be the set of feasible solutions of (2.1). Let $\tilde{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$ be two distinct BFSs of S, and $J = \{j : j$ such that either $\tilde{x}_j$ or $\widehat{x}_j$ or both are $> 0\}$. Let $L = \{\boldsymbol{x}(\alpha) = \alpha\tilde{\boldsymbol{x}} + (1-\alpha)\widehat{\boldsymbol{x}},\ 0 \leq \alpha \leq 1\}$. The following statements are equivalent:

1. If a point in L can be expressed as a convex combination of $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$, both $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are themselves in L.
2. The rank of $\{\boldsymbol{a}_j : j \in J\}$ is one less than its cardinality.

The algebraic characterization of adjacency provides a convenient method for checking whether a given pair of BFSs of (2.1) is adjacent. Now suppose $\boldsymbol{B} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$ is the current basis with corresponding BFS $\tilde{\boldsymbol{x}}$. Therefore, $\tilde{x}_{m+1} = \tilde{x}_{m+2} = \cdots = \tilde{x}_n = 0$. Let the entering variable in this step of the simplex algorithm be $\tilde{x}_{m+1}$ with pivot column $\boldsymbol{\alpha}_{m+1} = \boldsymbol{B}^{-1}\boldsymbol{a}_{m+1} = (\alpha_{1,m+1}, \ldots, \alpha_{m,m+1})^T$. We assume that $\boldsymbol{\alpha}_{m+1} \not\leq \boldsymbol{0}$. The case where $\boldsymbol{\alpha}_{m+1} \leq \boldsymbol{0}$ corresponds to unboundedness and is taken up in Section 2.3.3.

During the pivot step, the remaining nonbasic variables $x_{m+2}, \ldots, x_n$ are fixed at 0. Only the nonbasic variable x_{m+1} is changed from its present value of 0 to a

nonnegative δ and the values of the basic variables are modified to assure that the equality constraints are satisfied. From the discussion in Section 2.2.3 this leads to the solution $\boldsymbol{x}(\delta) = (\tilde{x}_1 - \alpha_{1,m+1}\delta, \ldots, \tilde{x}_m - \alpha_{m,m+1}\delta, \delta, 0, \ldots, 0)$. The maximum value that δ can take in this pivot step is the minimum ratio θ which is determined to assure that $\boldsymbol{x}(\delta)$ remains nonnegative. If the minimum ratio θ is zero, the new BFS is $\tilde{\boldsymbol{x}}$ itself. In this case the pivot step is degenerate and we have the following result.

1. In a degenerate pivot step, the simplex algorithm remains at the same BFS, but it obtains a new basis associated with it.

If the minimum ratio θ in this pivot step is strictly positive, the simplex algorithm moves to the new BFS $\boldsymbol{x}(\theta)$, which is distinct from $\tilde{\boldsymbol{x}}$. Clearly, the line segment joining $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}(\theta)$ is itself generated by varying δ in $\boldsymbol{x}(\delta)$ from 0 to θ. If $\bar{\boldsymbol{x}}$ is a point on this line segment then $\bar{x}_{m+2} = \cdots = \bar{x}_n = 0$. From Definition 2.3.2 we can then argue that if $\bar{\boldsymbol{x}} = \alpha\boldsymbol{x}^1 + (1-\alpha)\boldsymbol{x}^2$, where $0 < \alpha < 1$ and $\boldsymbol{x}^1, \boldsymbol{x}^2$ are feasible to (2.1) both $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ must be points of the form $\boldsymbol{x}(\delta)$. This implies that $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}(\theta)$ are themselves adjacent extreme points of S which, as a direct consequence of Theorem 2.3.2, leads to the following result.

2. During a nondegenerate pivot, the simplex algorithm moves from one extreme point $\tilde{\boldsymbol{x}}$ of the feasible set S to an adjacent extreme point $\boldsymbol{x}(\theta)$. The edge joining the extreme points is generated by giving the entering variable, call it $\tilde{x}_{m+1}$, all possible values between 0 and the minimum ratio θ. The equation of the corresponding edge is given by

$$\begin{aligned} x_i &= \tilde{x}_i - \alpha_{i,m+1}\delta, \quad i = 1, \ldots, m \\ x_{m+1} &= \delta \end{aligned}$$

for $0 \leqq \delta \leqq \theta$.

2.3.3 Unboundedness

Although most real problems are bounded, when using decomposition methods to solve large-scale applications, the subproblems that result are often unbounded. Therefore, it is worth investigating the geometry associated with this condition. We do so from the point of view presented in Chapter 3 of [M18]. The omitted proofs can be found therein.

Consider the system of equations in nonnegative variables in (2.1b). A homogeneous solution corresponding to this system is a vector $\boldsymbol{y} \in R^n$ satisfying

$$\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}, \quad \boldsymbol{y} \geq \boldsymbol{0}$$

The set of all homogeneous solutions is a convex polyhedral cone. If $\tilde{\boldsymbol{x}}$ is feasible to (2.1) and $\tilde{\boldsymbol{y}}$ is a corresponding homogeneous solution, then $\tilde{\boldsymbol{x}} + \delta\tilde{\boldsymbol{y}}$ is also a feasible solution to (2.1) for any $\delta \geqq 0$.

Theorem 2.3.3 *(Resolution Theorem I)* Every feasible solution of (2.1) can be expressed as the sum of (i) a convex combination of BFSs of (2.1) and (ii) a homogeneous solution corresponding to this system.

Proof: Suppose $\tilde{\boldsymbol{x}}$ is a feasible solution of (2.1). Let $\{j : \tilde{x}_j > 0\} = \{j_1, \ldots, j_k\}$ which means that the set of column vectors of $\boldsymbol{A}$ used by $\tilde{\boldsymbol{x}}$ is $\{\boldsymbol{a}_{j_1}, \ldots, \boldsymbol{a}_{j_k}\}$. The proof is based upon induction on the number of k, the number of columns vectors in this set.

Case 1: Suppose $k = 0$. This can only happen when $\tilde{\boldsymbol{x}} = \boldsymbol{0}$, and since $\tilde{\boldsymbol{x}}$ is assumed to be feasible, $\boldsymbol{b} = \boldsymbol{0}$. Thus $\tilde{\boldsymbol{x}} = \boldsymbol{0}$ is itself a BFS of (2.1) and also a homogeneous solution. Therefore

$$\tilde{\boldsymbol{x}} = \underbrace{\boldsymbol{0}}_{\text{BFS}} + \underbrace{\boldsymbol{0}}_{\substack{\text{Homogeneous} \\ \text{solution}}}$$

and the theorem holds.

Case 2: $k \geqq 0$.

Induction Hypothesis: Suppose the theorem holds for every feasible solution that uses a set of $k - 1$ or less column vectors of $\boldsymbol{A}$.

It must be shown that under this hypothesis, the theorem also holds for $\tilde{\boldsymbol{x}}$ which uses a set of k column vectors of $\boldsymbol{A}$. If $\{\boldsymbol{a}_{j_1}, \ldots, \boldsymbol{a}_{j_k}\}$ is linearly independent then $\tilde{\boldsymbol{x}}$ is a BFS of (2.1) so

$$\tilde{\boldsymbol{x}} = \underbrace{\tilde{\boldsymbol{x}}}_{\text{BFS}} + \underbrace{\boldsymbol{0}}_{\substack{\text{Homogeneous} \\ \text{solution}}}$$

and the result follows. For the case where the set $\{\boldsymbol{a}_{j_1}, \ldots, \boldsymbol{a}_{j_k}\}$ is not linearly independent, the reader is referred to [M18]. ∎

Boundedness of Convex Polytopes: If the set of feasible solutions $S = \{\boldsymbol{x} \in R^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqq \boldsymbol{0}\}$ is nonempty, by Resolution Theorem I, it is a convex polyhedron (i.e., bounded) iff $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$, $\boldsymbol{y} \geqq \boldsymbol{0}$ has a unique solution; namely, $\boldsymbol{y} = \boldsymbol{0}$.

Extreme Homogeneous Solutions: A homogeneous solution corresponding to (2.1) is called an *extreme homogeneous solution* iff it is a BFS of

$$\begin{aligned} \boldsymbol{A}\boldsymbol{y} &= \boldsymbol{0}, \quad \boldsymbol{y} \geqq \boldsymbol{0} \\ \sum_{j=1}^{n} y_j &= 1 \end{aligned} \tag{2.31}$$

Thus there are only a finite number of distinct extreme homogeneous solutions. The summation constraint in (2.31) normalizes the solution and eliminates $\mathbf{0}$ from consideration.

Lemma 2.3.1 Either $\mathbf{0}$ is a the unique homogeneous solution or every homogeneous solution corresponding to (2.1) can be expressed as a nonnegative linear combination of extreme homogeneous solutions.

Theorem 2.3.4 *(Resolution Theorem II for Convex Polytopes)* Let S be the set of feasible solutions to (2.1).

1. If $\mathbf{0}$ is the unique homogeneous solution corresponding to (2.1), every feasible solution can be expressed as a convex combination of BFSs of (2.1).
2. If $\mathbf{0}$ is not the only homogeneous solution corresponding to (2.1), every feasible solution is the sum of a convex combination of BFSs of (2.1) and a nonnegative combination of extreme homogeneous solutions corresponding to (2.1).

Proof: Follows from Resolution Theorem I and Lemma 2.3.1. ∎

Corollary 2.3.1 If the LP (2.1) has a feasible solution, it has an optimal feasible solution iff $f(\boldsymbol{y}) \geqq 0$ for every homogeneous solution y corresponding to (2.1).

Corollary 2.3.2 Suppose there exists an extreme homogeneous solution $\boldsymbol{y}^k$ corresponding to (2.1) such that $\boldsymbol{c}\boldsymbol{y}^k < 0$. If (2.1) is feasible, $f(\boldsymbol{x})$ is unbounded below on the set S of feasible solutions.

Corollary 2.3.3 If $f(\boldsymbol{x})$ is unbounded below on the set of feasible solutions of the LP (2.1), it remains unbounded for all values of the vector $\boldsymbol{b}$ that preserve feasibility.

Proof: Because $f(\boldsymbol{x})$ is unbounded below on S, by Corollaries 2.3.1 and 2.3.2 there exists a $\boldsymbol{y}^1$ satisfying $\boldsymbol{A}\boldsymbol{y}^1 = \mathbf{0}$, $\boldsymbol{c}\boldsymbol{y}^1 < 0$, $\boldsymbol{y}^1 \geqq \mathbf{0}$. Suppose $\boldsymbol{b}$ is changed to $\boldsymbol{b}^1$; let $\boldsymbol{x}^1$ be a feasible solution to the modified LP. Then $\boldsymbol{x}^1 + \delta\boldsymbol{y}^1$ is also feasible for all $\delta \geqq 0$, and since $\boldsymbol{c}(\boldsymbol{x}^1 + \delta\boldsymbol{y}^1) = \boldsymbol{c}\boldsymbol{x}^1 + \delta(\boldsymbol{c}\boldsymbol{y}^1)$ tends to $-\infty$ as δ tends to $+\infty$ (because $\boldsymbol{c}\boldsymbol{y}^1 < 0$), $f(\boldsymbol{x})$ is unbounded below as well in this modified problem. ∎

Corollary 2.3.4 Unboundedness of the set of feasible solutions: Let S denote the set of feasible solutions of (2.1) and let S^0 denote the set of homogeneous solutions corresponding to this system. The set S is unbounded iff $S \neq \emptyset$ and S^0 contains a nonzero point (i.e., the system (2.31) has a nonzero solution). Conversely, if $S^0 = \{\mathbf{0}\}$, S is bounded.

Proof: Follows from Theorem 2.3.4. ∎

Corollary 2.3.5 When (2.1) has a feasible solution, a necessary and sufficient condition for it to have a finite optimal feasible solution is that the optimal objective function value in the LP: min $\{\boldsymbol{c}\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \mathbf{0},\ \boldsymbol{x} \geqq \mathbf{0}\}$ is zero.

Proof: Follows from Corollary 2.3.1. ∎

2.3.4 Finding Adjacent Extreme Points

For the LP (2.1), let $\tilde{\boldsymbol{x}}$ be an extreme point of S. In a variety of applications as well as in some algorithms for the linear bilevel programming problem, it may be necessary to do one or more of the following.

1. Generate all adjacent extreme points of $\tilde{\boldsymbol{x}}$ on S.
2. Generate all adjacent extreme points of $\tilde{\boldsymbol{x}}$ on S such that $f(\boldsymbol{x}) \geqq f(\tilde{\boldsymbol{x}})$.
3. Generate all the bounded edges and unbounded edges of S containing $\tilde{\boldsymbol{x}}$.

When $\tilde{\boldsymbol{x}}$ is a nondegenerate BFS the computations for these three cases are easy, as discussed below. On the other hand, when $\tilde{\boldsymbol{x}}$ is a degenerate BFS the amount of work and bookkeeping goes up substantially. For this case, see [M9].

Suppose $\tilde{\boldsymbol{x}}$ is a nondegenerate BFS. Then exactly m of the $\tilde{\boldsymbol{x}}$ are positive. Let $J = \{j_1, \ldots, j_m\} = \{j : j \text{ such that } \tilde{x}_j > 0\}$ in this case. Then $\boldsymbol{x}_B = (\tilde{x}_{j_1}, \ldots, \tilde{x}_{j_m})$ is the unique basic vector for (2.1) associated with the BFS $\tilde{\boldsymbol{x}}$. Now obtain the canonical tableau for (2.1) with respect to the basic vector $\boldsymbol{x}_B$ and denote the entries by α_{ij}, $\bar{b}_i$, $\bar{c}_j$, $\bar{f}$ as usual. Using this notation, $\bar{b}_i = \tilde{x}_i$ for $i = 1, \ldots, m$, and $\bar{f} = f(\tilde{\boldsymbol{x}})$. By the nondegeneracy of $\tilde{\boldsymbol{x}}$, $\bar{b}_i > 0$ for all i. For each $j \notin J$, compute $\theta_j = \min \{\bar{b}_i/\alpha_{ij} : i \text{ such that } \alpha_{ij} > 0\}$, or $+\infty$ if $\alpha_{ij} \leqq 0$ for all i. The parameter θ_j is the minimum ratio when x_j enters the basis. For each $s \notin J$, define $\boldsymbol{x}^s(\delta) = (x_1^s(\delta), \ldots, x_n^s(\delta))$, where

$$x_{j_i}^s(\delta) = \bar{b}_i - \alpha_{is}\delta \quad \text{for } i = 1, \ldots, m$$

$$x_S^s(\delta) = \delta$$

$$x_j^s(\delta) = 0 \quad \text{for all } j \notin J \cup \{s\}$$

Then the set $\{\boldsymbol{x}^s(\theta_s) : s \notin J \text{ such that } \theta_s \text{ is finite}\}$ is the set of adjacent extreme points of $\tilde{\boldsymbol{x}}$ in S. The set of adjacent extreme points of $\tilde{\boldsymbol{x}}$ on S at which $f(\boldsymbol{x}) \geqq f(\tilde{\boldsymbol{x}})$ is $\{\boldsymbol{x}^s(\theta_s) : s \notin J \text{ such that } \bar{c}_s \geqq 0 \text{ and } \theta_s \text{ is finite}\}$. The set of unbounded edges of S through $\tilde{\boldsymbol{x}}$ is $\{\{\boldsymbol{x}^s(\delta) : \delta \geqq 0\} : s \notin J \text{ such that } \theta_s = +\infty\}$. The set of bounded edges of S through $\tilde{\boldsymbol{x}}$ is $\{\{\boldsymbol{x}^s(\delta) : \delta \geqq 0\} : s \notin J \text{ such that } \theta_s \text{ is finite}\}$.

2.3.5 Main Geometric Argument

Based on the discussions in the above sections, we can state the central mathematical result associated with the simplex method. Let S be a convex polytope in R^n and let $f(\boldsymbol{x})$ be a linear objective functional defined on it.

1. If $\boldsymbol{x}^*$ is any extreme point of S either $\boldsymbol{x}^*$ minimizes $f(\boldsymbol{x})$ on S or there exists an edge (bounded or unbounded) of S through $\boldsymbol{x}^*$ such that $f(\boldsymbol{x})$ decreases strictly as we move along this edge away from $\boldsymbol{x}^*$.

2. If $f(\boldsymbol{x})$ is bounded below on S and $\tilde{\boldsymbol{x}}$ is any extreme point of S, either $\tilde{\boldsymbol{x}}$ minimizes $f(\boldsymbol{x})$ on S or there exists an adjacent extreme point of $\tilde{\boldsymbol{x}}$ on S, say $\widehat{\boldsymbol{x}}$, such that $f(\widehat{\boldsymbol{x}}) < f(\tilde{\boldsymbol{x}})$.

For minimizing a linear functional $f(\boldsymbol{x})$ on a convex polytope S, the simplex algorithm starts at an extreme point of S and at each iteration moves along an edge of S incident to this extreme point, such that $f(\boldsymbol{x})$ decreases. Algorithms of this type are known as adjacent vertex methods. The class of optimization problems that can be solved by such methods include LPs, fractional linear programs, multiple objective linear programs, linear bilevel programs, quadratic programs, separable nonlinear programs, and a subclass of nonlinear programs whose objective functions satisfy certain monotonicity properties. In most of these cases, though, more efficient methods are available.

2.3.6 Alternative Optima and Uniqueness

Let $\boldsymbol{B}^*$ be an optimal BFS to the LP (2.1) with optimal objective function value f^*, and reduced cost coefficients $\bar{c}_j$. By the fundamental optimality criterion any feasible solution $\boldsymbol{x}$ that makes $f(\boldsymbol{x}) = f^*$ is also an optimal solution to this problem. Hence the set of optimal feasible solutions to (2.1) must satisfy

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqq \boldsymbol{0}$$
$$f(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x} = f^*$$

This leads to an alternative proof of Theorem 2.1.3 which can be rephrased as:

1. The set of optimal feasible solutions of an LP is itself a convex polytope, and
2. every convex combination of optimal feasible solutions is also optimal.

Because $\boldsymbol{B}^*$ is an optimal basis, all the reduced cost coefficients $\bar{c}_j$ must be nonnegative. Letting Q be the set of nonbasic variables, at optimality (2.27b) can be written as

$$f(\boldsymbol{x}) = f^* + \sum_{j \in Q} \bar{c}_j x_j$$

From this equation we see that every feasible solution to (2.1) in which the x_j variables with strictly positive $\bar{c}_j$'s are zero, is an optimal feasible solution. This is known as the *complementary slackness condition* and implies that if $T = \{j : j \text{ such that } \bar{c}_j > 0\}$, the set of optimal feasible solutions of (2.1) is the set of feasible solutions of (2.1) in which $x_j = 0$ for all $j \in T$.

This means that if x_q is nonbasic at the optimum and $\bar{c}_q = 0$, and if the operation of bringing x_q into the basis involves a nondegenerate pivot, doing so would yield an

alternative optimal BFS. This leads to the following result which is sufficient (but not necessary) for (2.1) to have a unique optimal solution.

Proposition 2.3.1 If there exists an optimal basis for the LP (2.1) in which the reduced cost coefficients of all the nonbasic variables are strictly positive, the corresponding BFS is the unique optimum.

2.3.7 Ranking Extreme Points

One of the first algorithms purposed for solving the linear bilevel programming problem is based on the idea of examining the vertices adjacent to the vertex at which the leader's objective function is independently minimized without regard to the follower's objective. The procedure parallels the idea of ranking the extreme points of a polytope and is discussed below.

Let S be the set of feasible solutions to the LP (2.1) which is assumed to have a finite optimum. Let E be the set of extreme points in S. When the simplex method is applied to this LP, it finds an optimal BFS, $\boldsymbol{x}^1 \in E$ at which $f(\boldsymbol{x})$ achieves its minimum value in S. Let $\boldsymbol{x}^2$ be an extreme point of S satisfying $f(\boldsymbol{x}^2) = \min\{f(\boldsymbol{x}) : \boldsymbol{x} \in E \setminus \{\boldsymbol{x}^1\}\}$. Then $\boldsymbol{x}^2$ is known as the second best extreme point solution for (2.1) or the second best BFS. It is possible that $f(\boldsymbol{x}^2) = f(\boldsymbol{x}^1)$, implying that $\boldsymbol{x}^2$ is an alternative optimum.

In ranking the extreme points of S in nondecreasing order of the value of $f(\boldsymbol{x})$ we are aiming for a sequence $\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \ldots$ with the property that

$$f(\boldsymbol{x}^{r+1}) = \min\{f(\boldsymbol{x}) : \boldsymbol{x} \in E \setminus \{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r\}\} \tag{2.32}$$

for each $r \geqq 1$. When determined in this manner, $\boldsymbol{x}^{r+1}$ is known at the $(r+1)$st best extreme point or BFS. Before presenting the algorithm we prove some theoretical results underlying its development.

Proposition 2.3.2 Let $\tilde{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$ be a pair of extreme points of S. Then there exists and edge path of S between $\tilde{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}}$.

Proof: For $\tilde{\boldsymbol{x}}$ a BFS of (2.1), let $J = \{j : \tilde{x}_j > 0\} = \{j_1, \ldots, j_k\}$ where $k \leqq m$. Define $d_j = 0$ if $j \in J$, 1 otherwise, and let $\psi(\boldsymbol{x}) = \sum_{j=1}^{n} d_j x_j$. Then $\psi(\tilde{\boldsymbol{x}}) = 0$ and for every $\boldsymbol{x} \in S$, $\boldsymbol{x} \neq \tilde{\boldsymbol{x}}$ we have $\psi(\boldsymbol{x}) > 0$. This means that $\tilde{\boldsymbol{x}}$ is the unique optimal solution of the LP: $\min\{\psi(\boldsymbol{x}) : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqq \boldsymbol{0}\}$. Apply the simplex algorithm to solve this problem starting with $\widehat{\boldsymbol{x}}$ as the initial BFS. Because $\tilde{\boldsymbol{x}}$ is the unique optimum of this LP, the algorithm walks along an edge path of S which must terminate at $\tilde{\boldsymbol{x}}$. ∎

Proposition 2.3.3 Let $\boldsymbol{x}^1$ be an optimal BFS of (2.1) and let $\tilde{\boldsymbol{x}}$ be another BFS. There exists an edge path of S from $\tilde{\boldsymbol{x}}$ to $\boldsymbol{x}^1$ with the property that as we walk along this path from $\tilde{\boldsymbol{x}}$ to $\boldsymbol{x}^1$, the value of $f(\boldsymbol{x})$ is nonincreasing.

Proof: If $\boldsymbol{c} = \boldsymbol{0}$, then all the feasible solutions of (2.1) have the same objective function value of zero and the result follows from Proposition 2.3.2. Assume now that $\boldsymbol{c} \neq \boldsymbol{0}$. Apply the simplex algorithm to solve (2.1) starting from the BFS $\tilde{\boldsymbol{x}}$. If $\boldsymbol{x}^1$ is the unique optimum the algorithm traces an edge path of S terminating at $\boldsymbol{x}^1$ along which $f(\boldsymbol{x})$ is nonincreasing. If $\boldsymbol{x}^1$ is not the unique optimum, it may terminate at an alternative optimal BFS $\widehat{\boldsymbol{x}}$. In this case the set of all optimal solutions is given by $F = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ f(\boldsymbol{x}^1) = \boldsymbol{c}\boldsymbol{x},\ \boldsymbol{x} \geqq \boldsymbol{0}\}$. Because F is a face of S, all extreme points and edges of F are similarly extreme points and edges of S. Now $\boldsymbol{x}^1$ and $\widehat{\boldsymbol{x}}$ are two extreme points of F. From Proposition 2.3.2 there exists an edge path in F from $\widehat{\boldsymbol{x}}$ to $\boldsymbol{x}^1$ and so every point on this path satisfies $f(\boldsymbol{x}) = f(\widehat{\boldsymbol{x}}) = f(\boldsymbol{x}^1)$. Combining the edge path from $\tilde{\boldsymbol{x}}$ to $\widehat{\boldsymbol{x}}$ obtained earlier during the simplex algorithm with the edge path from $\widehat{\boldsymbol{x}}$ to $\boldsymbol{x}^1$ in F, we have the required path. ∎

Proposition 2.3.4 Suppose that the r best extreme points of S in the ranked sequence, $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$, have already been determined. Then the $(r+1)$st best extreme point of S, $\boldsymbol{x}^{r+1}$, is adjacent to (and distinct from) one of these r points.

Proof: Let $\{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^t\}$ be the set of all adjacent extreme points of $\boldsymbol{x}^1$, or $\boldsymbol{x}^2$, ... or $\boldsymbol{x}^r$ on S excluding $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$. Let $\boldsymbol{x}^{r+1}$ be the point in $\{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^t\}$ that minimizes $f(\boldsymbol{x})$ among them. If $\widehat{\boldsymbol{x}}$ is any extreme point of S distinct from $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$, $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^t$, any edge path in S from $\boldsymbol{x}^1$ to $\widehat{\boldsymbol{x}}$ must contain one of the points $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^t$ as an intermediate point. Thus by Proposition 2.3.3, we must have $f(\widehat{\boldsymbol{x}}) \geqq f(\boldsymbol{x}^{r+1})$, implying that $\boldsymbol{x}^{r+1}$ determined in this manner satisfies (2.32). ∎

Ranking algorithm: If the r best extreme points of S have been determined, Proposition 2.3.4 says that is only necessary to look among the corresponding adjacent extreme points for the $(r+1)$st. This provides the main result for ranking the extreme points of S in nondecreasing order of the value of $f(\boldsymbol{x})$. The algorithm, taken from [M12], begins with an optimal BFS $\boldsymbol{x}^1$ of (2.1).

Step 1: Let $\boldsymbol{x}^1$ be the first element in the ranked sequence. Obtain all adjacent extreme points of $\boldsymbol{x}^1$ in S and store these in a `LIST` in increasing order of the objective function $f(\boldsymbol{x})$ from top to bottom. The procedure described in Section 2.3.4 can be used to generate the adjacent extreme points. While storing the extreme points in the list, it is convenient to store the corresponding basic vectors, their values, and the corresponding objective function values.

General Step: Suppose the extreme points $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$ in the ranked sequence have already been obtained. The list at this stage contains all the adjacent extreme points of $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$, excluding $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$, arranged in increasing order of their objective

function values from top to bottom. Select the extreme point at the top of the list and make it $\boldsymbol{x}^{r+1}$, the next extreme point in the ranked sequence, and remove it from the list. Obtain all the adjacent extreme points of $\boldsymbol{x}^{r+1}$ in S which satisfy $f(\boldsymbol{x}) \geqq f(\boldsymbol{x}^{r+1})$. If any one of them is not in the current list and not equal to any of the extreme points in the ranked sequence obtained so far, insert it in the appropriate slot in the list according to its objective function value. With this new list, continue until as many extreme points as required are obtained.

For large problems, the list of ranked order extreme points can get quite long since it is necessary to retain data on all extreme points adjacent to the first r. This is true even if we wish to obtain only the first k $(> r)$ best extreme points. The $(r+1)$st best extreme point may in fact be adjacent to any of the first r extreme points and not necessarily the rth. A final observation should be made about degeneracy. If the course of the algorithm, if a degenerate BFS appears in the ranked sequence, then finding all its adjacent extreme points requires additional bookkeeping and most likely an increase in computational complexity, as mentioned in Section 2.3.4.

2.4 ADDITIONAL FEATURES

In this section two extensions to the basic simplex method are discussed that are needed in practice. The first concerns the construction of an initial basic feasible solution; the second deals with the case where both upper and lower bounds are present on some of the variables. We conclude with a brief presentation of the linear complementarity problem and show that any linear program can be modeled equivalently. This result has led to the development of several algorithms for the linear bilevel programming problem, as we shall see in Chapter **??**.

2.4.1 Phase 1 and Artificial Variables

To start the simplex algorithm a basic feasible solution is required. Recall that the Fundamental Theorem 2.1.1 states in part that if an LP in standard form has a feasible solution it has a BFS. In some problems, a BFS can be found by inspection. Suppose $\boldsymbol{b} \geqq \boldsymbol{0}$ and $\boldsymbol{A}$ contains m unit vectors. Then the columns of $\boldsymbol{A}$ can be permuted to obtain the matrix $(\boldsymbol{A}', \boldsymbol{I}_m)$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix. Writing (2.1b) as $\boldsymbol{I}_m \boldsymbol{x}_B + \boldsymbol{A}' \boldsymbol{x}_N = \boldsymbol{b}$, it is clear that $\boldsymbol{x}_B = \boldsymbol{b}$, $\boldsymbol{x}_N = \boldsymbol{0}$ is a BFS. This case occurs in the situation where the original problem is

$$\begin{aligned} &\min \ f(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x} \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} \leqq \boldsymbol{b}, \ \boldsymbol{x} \geqq \boldsymbol{0} \end{aligned}$$

and m slack variables are added to each constraint giving

$$\begin{aligned} &\min\ f(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x} \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{I}_m\boldsymbol{x}_S = \boldsymbol{b},\ \boldsymbol{x} \geqq \boldsymbol{0},\ \boldsymbol{x}_S \geqq \boldsymbol{0} \end{aligned}$$

where $\boldsymbol{x}_S = \boldsymbol{x}_B$ and $\boldsymbol{x} = \boldsymbol{x}_N$.

Assume now that the constraints are written in standard form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \geqq \boldsymbol{0}$ and $\boldsymbol{A}$ does not contain an identity matrix. Without loss of generality, $\boldsymbol{A}$ can be partitioned as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{I}_k \\ \boldsymbol{A}_2 & \boldsymbol{0} \end{pmatrix}$$

where $\boldsymbol{I}_k$ is an identity matrix of order k ($0 \leqq k \leqq m$). Thus the constraints can be written as

$$\begin{aligned} \boldsymbol{A}_1\boldsymbol{x}_N + \boldsymbol{I}_k\boldsymbol{x}_k &= \boldsymbol{b}_1 \\ \boldsymbol{A}_2\boldsymbol{x}_N &= \boldsymbol{b}_2 \end{aligned} \tag{2.33}$$

where $\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)^T$. Now consider a related LP with an $(m-k)$-dimensional vector $\boldsymbol{x}_A \geqq \boldsymbol{0}$ of artificial variables and the constraint set

$$\begin{aligned} A_1\boldsymbol{x}_N + I_k\boldsymbol{x}_k &= \boldsymbol{b}_1 \\ A_2\boldsymbol{x}_N + I_{m-k}\boldsymbol{x}_A &= \boldsymbol{b}_2 \end{aligned} \tag{2.34}$$

These constraints have an obvious BFS $\boldsymbol{x}_k = \boldsymbol{b}_1$, $\boldsymbol{x}_A = \boldsymbol{b}_2$, and $\boldsymbol{x}_N = \boldsymbol{0}$. Moreover, (2.33) has a BFS iff (2.34) has a BFS with $\boldsymbol{x}_A = \boldsymbol{0}$. Beginning with this BFS, our goal is to find a BFS with $\boldsymbol{x}_A = \boldsymbol{0}$ or show that none exits. This is known as *phase 1* of the simplex method and can be achieved by solving the LP

$$\begin{aligned} \min\ \omega(\boldsymbol{x}) &= \boldsymbol{1}\boldsymbol{x}_A \\ \text{subject to}\ \ x_0 - \boldsymbol{c}_N\boldsymbol{x}_N - \boldsymbol{c}_k\boldsymbol{x}_k &= 0 \\ \boldsymbol{A}_1\boldsymbol{x}_N + \boldsymbol{I}_k\boldsymbol{x}_k &= \boldsymbol{b}_1 \\ \boldsymbol{A}_2\boldsymbol{x}_N + \boldsymbol{I}_{m-k}\boldsymbol{x}_A &= \boldsymbol{b}_2 \\ \boldsymbol{x}_N, \boldsymbol{x}_k, \boldsymbol{x}_A \geqq \boldsymbol{0},\ x_0 &\text{ unrestricted} \end{aligned} \tag{2.35}$$

where $\boldsymbol{1}$ is the row vector $(1, \ldots, 1)$ of appropriate length. The variable x_0 always remains basic. The initial BFS to (2.35) is $\boldsymbol{x}_k = \boldsymbol{b}_1 - A_1\boldsymbol{x}_N$, $\boldsymbol{x}_A = \boldsymbol{b}_2 - A_2\boldsymbol{x}_N$, $x_0 = \boldsymbol{c}_k\boldsymbol{b}_1 + (\boldsymbol{c}_N - \boldsymbol{c}_kA_1)\boldsymbol{x}_N$, $\boldsymbol{x}_N = \boldsymbol{0}$, and objective function value $\omega(\boldsymbol{x}) = \boldsymbol{1}(\boldsymbol{b}_2 - A_2\boldsymbol{x}_N)$.

For any feasible solution to (2.35), $\boldsymbol{x}_A > \boldsymbol{0}$ implies $\omega(\boldsymbol{x}) > 0$, and $\boldsymbol{x}_A = \boldsymbol{0}$ implies $\omega(\boldsymbol{x}) = 0$. Thus if an optimal solution to the phase 1 problem (2.35) is positive, (2.34) is infeasible as is the original problem. In contrast, any BFS to (2.35) with $\omega(\boldsymbol{x}) = 0$ provides a BFS to (2.34). This is obvious if such a BFS contains no artificial variables; for the case where the BFS contains artificial variables, it must be degenerate. Noting that $\boldsymbol{A}$ is assumed to have full row rank, by appealing to Corollary 2.1.1, we know that is possible to pivot in an equivalent number of original problem variables.

Once such a solution is found, the ω row, the artificial variables, and any original problem variables with positive reduced cost coefficients are dropped. (Actually, the artificial variables can be dropped when they leave the basis.) We then switch to *phase 2*, and continue with the minimization of x_0.

Example 2.4.1 Consider the following LP with $n = 5$, $m = 2$ written in standard form.

$$\begin{array}{rrrrrrrrrr} \min & 5x_1 & & & + & 21x_3 & & & & \\ \text{subject to} & x_1 & - & x_2 & + & 6x_3 & - & x_4 & = & 2 \\ & x_1 & + & x_2 & + & 2x_3 & & - x_5 & = & 1 \\ & & & & & x_1, \ldots, x_5 \geqq 0 & & & & \end{array}$$

To solve this problem with the two-phase method, we introduce two artificial variables x_6 and x_7 and phase 1 objective: $\min \omega = x_6 + x_7$. Rewriting the objective row by substituting for x_6 and x_7 gives

$$\begin{aligned} \omega &= 2 - x_1 + x_2 - 6x_3 + x_4 + 1 - x_1 - x_2 - 2x_3 + x_5 \\ &= 3 - 2x_1 - 8x_3 + x_4 + x_5 \end{aligned}$$

This leads to the first canonical tableau

		$\bar{\boldsymbol{b}}$	x_1	x_2	$\downarrow$ x_3	x_4	x_5	x_6	x_7
	x_0	2	5	0	21	0	0	0	0
$\leftarrow$	x_6	2	1	-1	(6)	-1	0	1	0
	x_7	1	1	1	2	0	-1	0	1
	$-\omega$	-3	-2	0	-8	1	1	0	0

Pivoting on the indicated element gives

		$\bar{\boldsymbol{b}}$	x_1	$\downarrow$ x_2	x_3	x_4	x_5	x_6	x_7
	x_0	-7	3/2	7/2	0	7/2	0	$-7/2$	0
	x_3	1/3	1	-1	1	-1	0	1/6	0
$\leftarrow$	x_7	1/3	2/3	(4/3)	0	1/3	-1	$-1/3$	1
	$-\omega$	$-1/3$	$-2/3$	$-4/3$	0	$-1/3$	1	4/3	0

A second pivot eliminates the artificial variables from the basis so they can be dropped along with the ω row. Switching to phase 2, we now minimize x_0.

		$\bar{\boldsymbol{b}}$	$\downarrow$ x_1	x_2	x_3	x_4	x_5
	x_0	$-63/8$	$-1/4$	0	0	21/8	21/8
	x_3	3/8	1/4	0	1	$-1/8$	$-1/8$
$\leftarrow$	x_2	1/4	(1/2)	1	0	1/4	$-3/4$

Pivoting out x_2 and pivoting in x_1 gives the next tableau which is optimal.

	$\bar{\boldsymbol{b}}$	x_1	x_2	x_3	x_4	x_5
x_0	$-31/4$	0	$1/2$	0	$11/4$	$9/4$
x_3	$1/4$	0	$-1/2$	1	$-1/4$	$1/4$
x_1	$1/2$	1	2	0	$1/2$	$-3/2$

This gives the final solution $f^* = -x_0 = 31/4$, $\boldsymbol{x}^* = (1/2, 0, 1/4, 0, 0, 0)$.

2.4.2 Bounded Variables

The most general case of the nonnegativity constraint (2.1c) $\boldsymbol{x} \geq \boldsymbol{0}$ restricts a particular variable x_j to be between two bounds; i.e., $l_j \leq x_j \leq u_j$. As mentioned in Section 2.1, it is always possible to redefine a decision variable so that its lower bound is zero. Recall this is done by replacing x_j with $\widehat{x}_j + l_j$ and its bounds with $0 \leqq \widehat{x}_j \leqq \widehat{u}_j = u_j - l_j$, $j = 1, \ldots, n$. To put the resultant formulation into standard form, it is necessary to add slack variables to the upper bound inequality giving $\widehat{x}_j + y_j = \widehat{u}_j$, where the n slack variables $y_j \geqq 0$. This leads to the following model

$$\min_{(\widehat{\boldsymbol{x}}, \boldsymbol{y})} \ f(\widehat{\boldsymbol{x}}) = \boldsymbol{c}\widehat{\boldsymbol{x}} \tag{2.36a}$$

$$\text{subject to } \boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{b} \tag{2.36b}$$

$$\widehat{\boldsymbol{x}} + \boldsymbol{y} = \widehat{\boldsymbol{u}} \tag{2.36c}$$

$$\widehat{\boldsymbol{x}} \geq \boldsymbol{0}, \ \boldsymbol{y} \geq \boldsymbol{0} \tag{2.36a}$$

Although the simplex method can be applied to problem (2.36), there is a high price to be paid in terms if computing and storage requirements. The addition of constraint (2.36c) has the effect of increasing the dimensions of the basis from an $m \times m$ matrix to an $(m+n) \times (m+n)$ matrix. This is extremely inefficient and, we will see below, unnecessary. In fact, upper bounds can be treated implicitly in the algorithm, in the same manner in which the lower bounds are treated.

To describe the augmented procedures, we introduce a new definition.

Definition 2.4.1 An *extended basic feasible solution* associated with an LP with variable bounds is a feasible solution for which $n - m$ variables are equal to either their lower (zero) or upper bound; the remaining m (basic) variables correspond to linearly independent columns of $\boldsymbol{A}$.

For convenience, we assume that every extended BFS is nondegenerate, which means that the m basic variables take values different from their bounds. Now suppose we start with an extended BFS and examine the nonbasic variables (the variables that are at one of their bounds) as possible candidates to enter the basis. A variable

at its lower bound can only be increased, and an increase will only be beneficial if its reduced cost coefficient is negative. A variable at its upper bound can only be decreased, and a decrease will only be beneficial if its reduced cost coefficient is positive. The value of nonbasic variables can be continuously changed until either (i) the value of a basic variable becomes equal to one of its bounds or (ii) the nonbasic variable being modified reaches its opposite bound. If (i) occurs the corresponding basic variable is declared nonbasic and the nonbasic variable takes its place in the basis. If (ii) occurs first, then the basis is not changed. The simultaneous occurrence of (i) and (ii) results in a degenerate solution which we have ruled out for now. The procedure continues until no further improvement is possible.

For notational simplicity, we consider a problem in which each of the variables has a lower bound of zero and arbitrary upper bound; i.e., $\min\{\boldsymbol{cx} : \boldsymbol{Ax} = \boldsymbol{b},\ \boldsymbol{0} \leqq \boldsymbol{x} \leqq \boldsymbol{u}\}$. Let Q denote the set of nonbasic variables, $Q_l = \{j : j \in Q,\ x_j = 0\}$ and $Q_u = \{j : j \in Q,\ x_j = u_j\}$. Modifying equations in (2.27) to reflect this partition gives

$$x_{B_i} = \bar{b}_i - \sum_{j \in Q_l} \alpha_{ij} x_j - \sum_{j \in Q_u} \alpha_{ij} x_j, \quad i = 1, \ldots, m$$

$$f = f^0 + \sum_{j \in Q_l} \bar{c}_j x_j + \sum_{j \in Q_u} \bar{c}_j x_j$$

with current values of x_{B_i} and f

$$x_{B_i} = \bar{b}_i - \sum_{j \in Q_u} \alpha_{ij} u_j, \quad i = 1, \ldots, m$$

$$f = f^0 + \sum_{j \in Q_u} \bar{c}_j u_j$$

Proceeding as in Section 2.2.3, it is easy to show that for $j \in Q_u$, $\bar{c}_j > 0$ implies that f can be decreased by moving x_j away from its upper bound u_j (perhaps making it basic). Analogous to Theorem 2.2.2, we have

Theorem 2.4.1 Given an extended basic feasible solution, a sufficient condition for optimality is

i) $0 \leqq x_{B_i} \leqq u_{B_i},\ i = 1, \ldots, m$ (feasibility)

ii) $\bar{c}_j \geqq 0,$ for all $j \in Q_l$

iii) $\bar{c}_j \leqq 0,$ for all $j \in Q_u$.

To incorporate these ideas into the simplex algorithm, we keep a record of whether a nonbasic variable is at zero (lower bound) or at its upper bound. In testing for optimality and choosing a variable to enter the basis, the opposite sign convention on $\bar{c}_j$ is used for a variable at its upper bounds ($\bar{c}_j \leqq 0$ for optimality; $\bar{c}_j > 0$ for

candidate to enter the basis). The rule for selecting a departing variable must also be modified to yield a first basic variable that reaches zero or its upper bound.

Suppose that $x_q,\ q \in Q_l$ is the entering variable. The first variable to reach zero (call it x_{B_p}), as discussed in Section 2.2.3, is determined from the minimum ratio test

$$\theta_{pq} = \min_{i=1,\ldots,m} \left\{ \frac{\bar{b}_i}{\alpha_{iq}} \;\middle|\; \alpha_{iq} > 0 \right\}$$

The first variable to reach its upper bound (excluding x_q) is given by x_{B_s} determined from

$$\beta_{sq} = \min_{i=1,\ldots,m} \left\{ \frac{\bar{b}_i - u_{B_i}}{\alpha_{iq}} \;\middle|\; \alpha_{iq} < 0 \right\}$$

Finally, it is necessary to maintain $x_q \leqq u_q$. Thus the departing variable is determined from

$$\min\{\theta_{pq}, \beta_{sq}, u_q\}$$

where ties are broken arbitrarily (the nondegeneracy assumption implies that the choice will be unique). If θ_{pq} or β_{sq} is minimum, x_{B_p} or x_{B_s}, respectively, departs and a simplex iteration is executed. If u_q is the minimum, x_q is both the entering and leaving variable so there is no change of basis. Nevertheless, the change of x_q from zero to u_q must be recorded and the solution vector and objective value must be updated as follows

$$\widehat{x}_{B_i} = \bar{b}_i - \alpha_{iq} u_q, \quad i = 1, \ldots, m$$

$$\widehat{f}^0 = f^0 + \bar{c}_q u_q$$

The case in which $x_q,\ q \in Q_u$ is the entering variable yields similar results. The first variable to reach zero (excluding x_q) is given by x_{B_p} determined from

$$\widehat{\theta}_{pq} = \max_{i=1,\ldots,m} \left\{ \frac{\bar{b}_i}{\alpha_{iq}} \;\middle|\; \alpha_{iq} < 0 \right\}$$

Here $\widehat{\theta}_{pq} \leqq 0$ represents the change allowed in x_q, so if $u_q < -\widehat{\theta}_{pq}$, x_q reaches zero before x_{B_p}. The first variable to reach its upper bound is x_{B_s} determined from

$$\widehat{\beta}_{sq} = \min_{i=1,\ldots,m} \left\{ \frac{\bar{b}_i - u_{B_i}}{\alpha_{iq}} \;\middle|\; \alpha_{iq} > 0 \right\}$$

In this case, the departing variable is determined from

$$\min\{\widehat{\theta}_{pq}, \widehat{\beta}_{sq}, -u_q\}$$

Algorithm Modifications

When bounded variables are present, the computations in the simplex algorithm outlined in Section 2.2.3 proceed as usual except that the choice of the pivot element

must be modified slightly. To describe the modifications, it is convenient to introduce the notation $x_j^+ = x_j$, $x_j^- = u_j - x_j$, which allows us to maintain the optimality condition $\bar{c}_j \geqq 0$ for all j nonbasic. As the method progresses we change back and forth from x_j^+ to x_j^-, depending on whether the variable x_j has most recently been at its lower or upper bound, respectively.

Using this notation, the procedure for moving from one extended BFS to another can be easily implemented by following the strategy outlined above.

Step 1. Determine a nonbasic variable x_q^+ for which $\bar{c}_q < 0$. If no such variable exists, stop; the current solution is optimal.

Step 2. Evaluate the three terms

a) u_q (upper bound on x_q)

b) $\min \left\{ \dfrac{\bar{b}_i}{\alpha_{iq}} \;\middle|\; \alpha_{iq} > 0 \right\}$

c) $\min \left\{ \dfrac{\bar{b}_i - u_q}{\alpha_{iq}} \;\middle|\; \alpha_{iq} < 0 \right\}$

Step 3. According to which number in Step 2 is smallest, update the basis inverse $\boldsymbol{B}^{-1}$ and the current solution $\boldsymbol{B}^{-1}\boldsymbol{b}$ at Step 4 of the revised simplex method as follows.

a) The variable x_q goes to its opposite bound. Subtract u_q times updated column q ($\boldsymbol{\alpha}_q$) from $\boldsymbol{B}^{-1}\boldsymbol{b}$. Multiply column q as well as $\bar{c}_q$ by minus one and note a change in the sign of the variable. The basis doesn't change and no pivot is required.

b) Suppose p is the minimizing index in (b) of Step 2. Then the pth basic variable returns to its old bound and we pivot on the pqth element α_{pq}.

c) Suppose p is the minimizing index in (c) of Step 2. Then the pth basic variable goes to its opposite bound. Subtract u_p from $\bar{b}_p$, change the sign of α_{pp}, pivot on the pqth element, and note a change in the sign of variable p.

Return to Step 1.

2.4.3 Kuhn-Tucker Conditions and the Linear Complementarity Problem

Let us consider the LP (2.1) in slightly different form: $\min\{\boldsymbol{cx} : \boldsymbol{Ax} \geqq \boldsymbol{b},\ \boldsymbol{x} \geqq \boldsymbol{0}\}$ and define the corresponding Lagrangian as in Section **??**.

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{cx} - \boldsymbol{\lambda}^T(\boldsymbol{Ax} - \boldsymbol{b}) - \boldsymbol{\mu}^T\boldsymbol{x}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the m- and n-dimensional Kuhn-Tucker or Lagrange multipliers for the technological and nonnegativity constraints, respectively. Writing the Kuhn-Tucker necessary conditions (see Theorem **??**) for optimality gives

$$\begin{aligned}
&\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c} - \boldsymbol{\lambda}^T \boldsymbol{A} - \boldsymbol{\mu}^T = \boldsymbol{0} \\
&\nabla_{\lambda} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \geqq \boldsymbol{0} \\
&\nabla_{\mu} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{x} \geqq \boldsymbol{0} \\
&\boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = 0 \\
&\boldsymbol{\mu}^T \boldsymbol{x} = 0 \\
&\boldsymbol{\lambda} \geqq \boldsymbol{0},\ \boldsymbol{\mu} \geqq \boldsymbol{0}
\end{aligned}$$

which, due to the convexity of the linear program, are sufficient conditions as well. This system can be expressed equivalently as

$$\begin{aligned}
&\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \geqq \boldsymbol{0}, \quad \boldsymbol{x} \geqq \boldsymbol{0} \\
&\boldsymbol{\lambda}^T \boldsymbol{A} \leqq \boldsymbol{c}, \qquad \boldsymbol{\lambda} \geqq \boldsymbol{0} \\
&\boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = (\boldsymbol{c} - \boldsymbol{\lambda}^T \boldsymbol{A})\boldsymbol{x} = 0
\end{aligned} \tag{2.37}$$

which we show in the next section are exactly the primal and dual conditions for optimality. The last equation in (2.37) implies $\boldsymbol{\lambda}^T \boldsymbol{b} = \boldsymbol{c}\boldsymbol{x}$. Introducing slack variables $\boldsymbol{u}$ and $\boldsymbol{v}$, the system becomes

$$\boldsymbol{A}\boldsymbol{x} \qquad - \boldsymbol{v} \qquad = \boldsymbol{b} \tag{2.38a}$$

$$\boldsymbol{A}^T \boldsymbol{\lambda} \qquad + \boldsymbol{u} = \boldsymbol{c}^T \tag{2.38b}$$

$$\boldsymbol{c}\boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{\lambda} \qquad = 0 \tag{2.38c}$$

$$\boldsymbol{x} \geq \boldsymbol{0},\ \boldsymbol{\lambda} \geq \boldsymbol{0},\ \boldsymbol{u} \geq \boldsymbol{0},\ \boldsymbol{v} \geq \boldsymbol{0} \tag{2.38d}$$

Thus if $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{u}}, \widehat{\boldsymbol{v}})$ is a feasible solution to (2.38), then $\widehat{\boldsymbol{x}}$ is an optimal feasible solution to the LP defined with inequality constraints. Conversely, if $\widehat{\boldsymbol{x}}$ is an optimal solution to the LP with optimal basis $\widehat{\boldsymbol{B}}$, then for $\widehat{\boldsymbol{\lambda}} = \boldsymbol{c}_B \boldsymbol{B}^{-1}$ there exist vectors $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}$ that satisfy (2.38).

This simple analysis has shown that solving an LP is equivalent to solving the system of linear equations (2.38) in nonnegative variables. Hence, any algorithm for solving such systems can be used to solve LPs directly without any need for optimization (in fact, this is what interior point methods do). Conversely, any algorithm for solving LPs can be used to solve a system of linear equations in nonnegative variables by solving a phase-1-type problem.

Going one step farther and rewriting (2.38a)–(2.38b) as $\boldsymbol{v} = A\boldsymbol{x} - \boldsymbol{b}$, $\boldsymbol{u} = A^T \boldsymbol{\lambda} - \boldsymbol{c}^T$, and noting that (2.37) implies $\boldsymbol{u}^T \boldsymbol{x} + \boldsymbol{v}^T \boldsymbol{\lambda} = 0$, we get a slightly different representa-

tion of the optimality conditions. Now let

$$\boldsymbol{\zeta} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} \quad \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} \quad \boldsymbol{M} = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \quad \boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{c}^T \\ -\boldsymbol{b} \end{pmatrix}$$

Solving a linear program written with inequality constraints is equivalent to solving the system

$$\begin{aligned} &\boldsymbol{\zeta} - \boldsymbol{M}\boldsymbol{\eta} = \boldsymbol{\nu} \\ &\boldsymbol{\zeta} \geqq \boldsymbol{0},\ \boldsymbol{\eta} \geqq \boldsymbol{0} \\ &\boldsymbol{\zeta}^T \boldsymbol{\eta} = 0 \end{aligned} \tag{2.39}$$

The constraints $\boldsymbol{\zeta} \geqq \boldsymbol{0}$, $\boldsymbol{\eta} \geqq \boldsymbol{0}$, $\boldsymbol{\zeta}^T\boldsymbol{\eta} = 0$ imply that $\zeta_j \eta_j = 0$ for all j. Hence at most one variable in each complementary pair (ζ_j, η_j) can take a positive value in any solution. Problems of the form (2.39) are known as linear complementarity problems and have a number of applications in optimization.

2.5 DUALITY

In this section, an elegant symmetrical relationship is derived for the linear programming problem. It will be shown that associated with every LP is a corresponding dual LP. Both problems are constructed from the same underlying cost and constraint data but in such a way that if one is a minimization problem the other is a maximization problem, and the optimal values of either objective function, if finite, are equal. The variables of the dual problem can be interpreted as prices associated with the constraints of the original or primal problem, and through this association it is possible to give an economic characterization of the dual whenever there is such a characterization for the primal.

Initially, we depart from the standard form of the LP (2.1) to highlight the symmetry of the relationship. Specifically, the following pair of LPs denoted by (P) and (D), respectively, are considered.

Primal		*Dual*	
min f =	$\boldsymbol{cx}$	max ϕ =	$\boldsymbol{\lambda}^T\boldsymbol{b}$
subject to	$\boldsymbol{Ax} \geqq \boldsymbol{b}$	subject to	$\boldsymbol{\lambda}^T\boldsymbol{A} \leq \boldsymbol{c}^T$
	$\boldsymbol{x} \geqq \boldsymbol{0}$		$\boldsymbol{\lambda} \geqq \boldsymbol{0}$

(2.40)

where $\boldsymbol{x} \in R^n$ and $\boldsymbol{\lambda} \in R^m$ are the primal and dual variables, respectively, and the coefficients are as previously defined.

Example 2.4.1 with the surplus variables removed is

$$\begin{array}{lrrrrr} \min\ f\ = & 5x_1 & & + 21x_3 & & \\ \text{subject to} & x_1 & - x_2 & + 6x_3 & \geqq & 2 \\ & x_1 & + x_2 & + 2x_3 & \geqq & 1 \\ & & & x_1, x_2, x_3 & \geqq & 0 \end{array}$$

The dual of this problem is

$$\begin{array}{lrrrr} \max\ \phi\ = & 2\lambda_1 & + \lambda_2 & & \\ \text{subject to} & \lambda_1 & + \lambda_2 & \leqq & 5 \\ & -\lambda_1 & + \lambda_2 & \leqq & 0 \\ & 6\lambda_1 & + 2\lambda_2 & \leqq & 21 \\ & & \lambda_1,\ \lambda_2 & \geq & 0 \end{array}$$

The optimal solution including slack variables is $\boldsymbol{\lambda}^* = (11/4, 9/4, 0, 1/2, 0)$, $f^* = 31/4$. The corresponding solution to the primal problem including surplus variables is $f^* = 31/4$, $\boldsymbol{x}^* = (1/2, 0, 1/4, 0, 0, 0)$. Note that the primal and dual objective values are equal. This is always the case.

The dual of any LP can be found by converting it to the form of the primal problem (P). For example, given an LP in standard form

$$\begin{array}{ll} \min\ f\ = & \boldsymbol{cx} \\ \text{subject to} & \boldsymbol{Ax} = \boldsymbol{b} \\ & \boldsymbol{x} \geqq \boldsymbol{0} \end{array}$$

we can replace the constraints $\boldsymbol{Ax} = \boldsymbol{b}$ as two inequalities: $\boldsymbol{Ax} \leqq \boldsymbol{b}$ and $-A\boldsymbol{x} \leqq -\boldsymbol{b}$ so the coefficient matrix becomes $\begin{pmatrix} \boldsymbol{A} \\ -\boldsymbol{A} \end{pmatrix}$ and the right-hand-side vector becomes $\begin{pmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \end{pmatrix}$. Introducing a partitioned dual row vector $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, the corresponding dual is

$$\begin{array}{ll} \max\ \phi\ = & \boldsymbol{\gamma}_1\boldsymbol{b} - \boldsymbol{\gamma}_2\boldsymbol{b} \\ \text{subject to} & \boldsymbol{\gamma}_1\boldsymbol{A} - \boldsymbol{\gamma}_2\boldsymbol{A} \leqq \boldsymbol{c}^T \\ & \boldsymbol{\gamma}_1 \geqq \boldsymbol{0}, \boldsymbol{\gamma}_2 \geqq \boldsymbol{0} \end{array}$$

Letting $\boldsymbol{\lambda}^T = \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2$ we may simplify the representation of this problem to obtain the following pair:

$$\begin{array}{ll} \textit{Primal} & \\ \min\ f\ = & \boldsymbol{cx} \\ \text{subject to} & \boldsymbol{Ax} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{array} \qquad \begin{array}{ll} \textit{Dual} & \\ \max\ \phi\ = & \boldsymbol{\lambda}^T\boldsymbol{b} \\ \text{subject to} & \boldsymbol{\lambda}^T\boldsymbol{A} \leq \boldsymbol{c}^T \end{array} \tag{2.41}$$

This is the asymmetric form of the duality relation. Similar transformations can be worked out for any linear program by first putting the primal in the form of

(2.40), constructing the dual, and then simplifying the latter to account for special structure. We say two LPs are equivalent if one can be transformed into another so that feasible solutions, optimal solutions and corresponding dual solutions are preserved; e.g., (2.40) and (2.41) are equivalent primal–dual representations.

2.5.1 Primal–Dual Relationship

Although it is always possible to transform an LP into the primal–dual pairs in either (2.40) or (2.41), it is often inconvenient to make this transformation for problems with a variety of constraints and variable types. Therefore, it is useful to state the rules for constructing a dual problem. To begin, there will be one dual variable associated with each constraint in the primal (excluding nonnegativity restrictions on individual variables, if any). If the primal constraint is a $\geqq$ inequality, the dual variable is restricted to being nonnegative, and vice versa. If the primal constraint is an equality, the dual variable is unrestricted in sign.

There is one dual constraint corresponding to each primal variable. Thus each column vector of the primal tableau leads to a constraint in the dual problem. Let $\boldsymbol{\lambda}^T$ be the row vector of dual variables. Let $\boldsymbol{a}_j$ be the column vector of the coefficients of the primal variable x_j and c_j its objective function coefficient. Then the dual constraint corresponding to x_j is $\boldsymbol{\lambda}^T \boldsymbol{a}_j = c_j$ if x_j is unrestricted in sign in the primal problem. If x_j is restricted to being nonnegative (nonpositive), the corresponding inequality in the dual problem is $\boldsymbol{\lambda}^T \boldsymbol{a}_j \leqq c_j$ ($\boldsymbol{\lambda}^T \boldsymbol{a}_j \geqq c_j$).

If the primal is a minimization problem, the dual is a maximization problem and vice versa. The right-hand-side constants of the primal are the coefficients of the dual objective function and vice versa. Any LP can be designated the primal whether or not it matches a standard structure.

2.5.2 Dual Theorems

To this point, the relation between the primal and dual problems has been simply a formal one based on what might appear as an arbitrary definition. The material in Section 2.4.3 relating to the Kuhn-Tucker conditions, though, suggests otherwise. In this section, a deeper connection between the pair is derived in terms of feasibility, optimality and bounds.

Proposition 2.5.1 Dual of the dual is primal.

Proof: For any primal–dual pair, this can be verified by writing out the corresponding dual and defining the variables appropriately. ∎

Proposition 2.5.2 Duals of equivalent problems are equivalent. Let (P) refer to an LP and let (D) be its dual. Let $(\widehat{\mathrm{P}})$ be an LP that is equivalent to (P). Let $(\widehat{\mathrm{D}})$ be the dual of $(\widehat{\mathrm{P}})$. Then $(\widehat{\mathrm{D}})$ is equivalent to (D); i.e., they have the same optimal objective function values or they are both infeasible.

Proof: By construction. ∎

Theorem 2.5.1 (*Weak Duality Theorem*) In a primal–dual pair of LPs, let $\boldsymbol{x}$ be a primal feasible solution and $f(\boldsymbol{x})$ the corresponding value of the primal objective function that is to be minimized. Let $\boldsymbol{\lambda}$ be a dual feasible solution and $\phi(\boldsymbol{\lambda})$ the corresponding dual objective function that is to be maximized. Then $f(\boldsymbol{x}) \geqq \phi(\boldsymbol{\lambda})$.

Proof: We consider the case where the primal and the dual are stated as in (2.40). Then

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &\geqq \boldsymbol{b} && \text{(because } \boldsymbol{x} \text{ is primal feasible)} \\ \boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{x} &\geqq \boldsymbol{\lambda}^T\boldsymbol{b} && \text{(because } \boldsymbol{\lambda} \geqq \boldsymbol{0}\text{)} \qquad (2.42) \\ \boldsymbol{\lambda}^T\boldsymbol{A} &\leqq \boldsymbol{c} && \text{(because } \boldsymbol{\lambda} \text{ is dual feasible)} \\ \boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{x} &\leqq \boldsymbol{c}\boldsymbol{x} && \text{(because } \boldsymbol{x} \geqq \boldsymbol{0}\text{)} \qquad (2.43) \end{aligned}$$

Combining (2.42) and (2.43) we get $\boldsymbol{c}\boldsymbol{x} \geq \boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{\lambda}^T\boldsymbol{b}$; that is, $f(\boldsymbol{x}) \geq \phi(\boldsymbol{\lambda})$, which proves the theorem when the primal and dual are stated in this form. In general, every LP can be transformed into an equivalent problem identical to the primal in (2.40). ∎

Corollaries of the Weak Duality Theorem

Considering any primal–dual pair of LPs, let the primal refer to the minimization problem and the dual to the maximization problem in the pair.

1. The primal objective value of any primal feasible solution is an upper bound to the maximum value of the dual objective in the dual problem.
2. The dual objective value of any dual feasible solution is a lower bound to the minimum value of the primal objective in the primal problem.
3. If the primal problem is feasible and its objective function is unbounded below on the primal feasible solution set, the dual problem cannot have a feasible solution.
4. If the dual problem is feasible and its objective function is unbounded above on the dual feasible solution set, the primal problem cannot have a feasible solution.
5. The converse of (3) is the following: If the dual problem is infeasible and the primal problem is feasible, the primal objective function is unbounded below on the primal feasible solution set. Similarly, the converse of (4) is: If the primal problem is infeasible and the dual problem is feasible, the dual objective function is unbounded above on the dual feasible solution set.

It is possible that both the primal and the dual problems in a primal–dual pair have no feasible solutions. For example, consider the following:

$$\begin{array}{lrcl} \min f = & 3x_1 - 5x_2 & & \\ \text{subject to} & x_1 - x_2 & = & 1 \\ & -x_1 + x_2 & = & 3 \\ & x_1 \geqq 0, x_2 \geqq 0 & & \end{array} \qquad \begin{array}{lrcl} \max \phi = & \lambda_1 + \lambda_2 & & \\ \text{subject to} & \lambda_1 - \lambda_2 & \leqq & 3 \\ & -\lambda_1 + \lambda_2 & \leqq & -5 \\ & \lambda_1, \lambda_2 \text{ unrestricted} & & \end{array}$$

Clearly both problems in the pair are infeasible. Thus even though the result in (5) is true, the fact that the dual problem is infeasible in a primal–dual pair of LPs does not imply that the primal objective function is unbounded on the primal feasible solution set, unless it is known that the primal is feasible.

Theorem 2.5.2 (*Sufficient Optimality Criterion*) In a primal–dual pair of LPs, let $\phi(\boldsymbol{x})$ be the primal objective function and $f(\boldsymbol{\lambda})$ be the dual objective function. If $\widehat{\boldsymbol{x}}$, $\widehat{\boldsymbol{\lambda}}$ are a pair of primal and dual feasible solutions satisfying $f(\widehat{\boldsymbol{x}}) = \phi(\widehat{\boldsymbol{\lambda}})$, then $\widehat{\boldsymbol{x}}$ is an optimal feasible solution of the primal and $\widehat{\boldsymbol{\lambda}}$ is an optimum feasible solution of the dual.

Proof: Suppose the primal denotes the minimization problem in the primal–dual pair. Let $\boldsymbol{x}$ be any primal feasible solution. By the weak duality theorem, we have $f(\boldsymbol{x}) \geqq \phi(\widehat{\boldsymbol{\lambda}})$ because $\widehat{\boldsymbol{\lambda}}$ is dual feasible. But $f(\widehat{\boldsymbol{x}}) = \phi(\widehat{\boldsymbol{\lambda}})$ by hypothesis. So $f(\boldsymbol{x}) \geqq f(\widehat{\boldsymbol{x}})$ for all $\boldsymbol{x}$ primal feasible. Thus $\widehat{\boldsymbol{x}}$ is optimal to the primal problem. Similarly, $\widehat{\boldsymbol{\lambda}}$ is optimal to the dual problem. ∎

Theorem 2.5.3 (*Fundamental Duality Theorem*) In a primal–dual pair of LPs, if either the primal or the dual problem has an optimal feasible solution, then the other does also and the two optimal objective values are equal.

Proof: We will prove the result for the case where the primal and dual problems are stated as in (2.41). Solving the primal problem by the simplex algorithm in Section 2.2.3 yields an optimal solution $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$, $\boldsymbol{x}_N = \boldsymbol{0}$ with $\bar{\boldsymbol{c}} = \boldsymbol{c}_N - \boldsymbol{c}_B\boldsymbol{B}^{-1}N \geqq \boldsymbol{0}$, which can be written $[\boldsymbol{c}_B, \boldsymbol{c}_N - \boldsymbol{c}_B\boldsymbol{B}^{-1}(\boldsymbol{B}, \boldsymbol{N})] = \boldsymbol{c} - \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqq \boldsymbol{0}$. Now if we define $\boldsymbol{\lambda}^T = \boldsymbol{c}_B\boldsymbol{B}^{-1}$ we have $\boldsymbol{\lambda}^T A \leqq \boldsymbol{c}$ and $f(\boldsymbol{x}) = \boldsymbol{c}_B\boldsymbol{x}_B = \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b} = \boldsymbol{\lambda}^T\boldsymbol{b} = \phi(\boldsymbol{\lambda})$. By the sufficient optimality criterion, Theorem 2.5.2, λ is a dual optimal solution. This completes the proof when the primal and dual are as stated here.

In general, every LP can be transformed into an equivalent problem in standard form. This equivalent problem is of the same type as the primal in (2.41), hence the proof applies. Also, by Proposition 2.5.2, the dual of the equivalent problem in standard form is equivalent to the dual of the original problem. Thus the fundamental duality theorem must hold for it too. ∎

Corollaries of the Fundamental Duality Theorem

Alternative statement of the duality theorem: "If both problems in a primal–dual pair of LPs have feasible solutions, then both have optimal feasible solutions and the

optimal objective values of the two problems are equal." This is easily proved by using the weak duality theorem and the fundamental duality theorem.

Separation property of objective values: Consider a primal–dual pair of LPs and suppose the minimization problem in the pair is the primal with the objective function $f(\boldsymbol{x})$. Suppose the dual objective function is $\phi(\boldsymbol{\lambda})$. If both problems have feasible solutions, then the values assumed by the two objective functions at feasible solutions of the respective problems are separated on the real line; i.e., $\phi(\boldsymbol{\lambda}) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$ and $\boldsymbol{\lambda}$.

Primal objective unbounded: If the primal is the minimization problem in a primal–dual pair, and if the primal is feasible and the dual is infeasible, then the primal cannot have an optimal feasible solution, that is, the primal objective function is unbounded below.

Dual objective unbounded: If the dual is the maximization problem in a primal–dual pair, and if the dual is feasible and the primal infeasible, then the dual cannot have an optimal feasible solution, that is, the dual objective function is unbounded above.

Proposition 2.5.3 (*Necessary and Sufficient Optimality Conditions*) Consider a primal–dual pair of LPs. Let $\boldsymbol{x}$, $\boldsymbol{\lambda}$ be the vectors of primal and dual variables and let $f(\boldsymbol{x})$, $\phi(\boldsymbol{\lambda})$ be the primal and dual objective functions, respectively. If $\widehat{\boldsymbol{x}}$ is a primal feasible solution, it is an optimal solution of the primal problem iff there exists a dual feasible solution $\widehat{\boldsymbol{\lambda}}$ satisfying $f(\boldsymbol{x}) = \phi(\boldsymbol{\lambda})$.

Proof: Follows directly from Theorems 2.5.3 and 2.5.2. ∎

Theorem 2.5.4 (*Complementary Slackness Theorem*) A pair of primal and dual feasible solutions are optimal to the respective problems in a primal–dual pair of LPs iff whenever these feasible solutions make a slack variable in one problem strictly positive, the value (in these feasible solutions) of the associated nonnegative variable of the other problem is zero.

For the primal–dual pair (2.40) the theorem has the following interpretation. Whenever

$$v_i = \sum_{j=1}^{n} a_{ij}x_j - b_i > 0 \quad \text{we have } \lambda_i = 0 \tag{2.44a}$$

$$u_j = c_j - \sum_{i=1}^{m} a_{ij}\lambda_i > 0 \quad \text{we have } x_j = 0 \tag{2.44b}$$

Alternatively, we have

$$v_i\lambda_i = \left(\sum_{j=1}^{n} a_{ij}x_j - b_i\right)\lambda_i = 0, \quad i = 1, \ldots, m \tag{2.45a}$$

$$u_j x_j = \left(c_j - \sum_{i=1}^{m} a_{ij}\lambda_i \right) x_j = 0, \quad j = 1, \ldots, n \tag{2.45b}$$

Remark 1 Conditions (2.44) or (2.45) only require that if $v_i > 0$, then $\lambda_i = 0$. They do not require that if $v_i = 0$, then λ_i must be > 0; that is, both v_i and λ_i could be zero and the conditions of the complementary slackness theorem would be satisfied. Moreover, conditions (2.44) and (2.45) automatically imply that if $\lambda_i > 0$, then $v_i = 0$, and that if $x_j > 0$, then $u_j = 0$.

Remark 2 The complementary slackness theorem does not say anything about the values of unrestricted variables (corresponding to equality constraints in the other problem) in a pair of optimal feasible solutions of the primal and dual problems, respectively. It is concerned only with nonnegative variables of one problem and the slack variables corresponding to the associated inequality constraints in the other problem of primal–dual LPs.

Corollary 2.5.1 Consider a primal–dual pair of LPs. Let $\boldsymbol{x}^*$ be an optimal feasible solution of the primal LP. Then the following statements can be made about every dual optimum feasible solution.

1. If x_j is a variable restricted to be nonnegative in the primal problem and if $x_j^* > 0$, then the dual inequality constraint associated with the primal variable x_j is satisfied as an equation by every dual optimal feasible solution.
2. If the primal problem consists of any inequality constraints, let $\boldsymbol{v}^*$ represent the values of the corresponding slack variables at the primal feasible solution $\boldsymbol{x}$. Then if a slack variable $v_i > 0$, the dual variable associated with it is equal to zero in every dual optimal feasible solution.

Corresponding symmetric statements can be made about the nonnegatively restricted dual variables in $\boldsymbol{\lambda}^*$ when they take positive values, and primal inequality constraints that must hold as equalities in every primal optimal solution.

Proposition 2.5.4 (*Necessary and Sufficient Optimality Conditions*) Let $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ be the vectors of variables in an LP and its dual, respectively. If $\widehat{\boldsymbol{x}}$ is a primal feasible solution it is an optimal solution iff there exists a dual feasible solution $\widehat{\boldsymbol{\lambda}}$ such that $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{\lambda}}$ together satisfy the complementary slackness conditions for optimality in this primal–dual pair.

Proof: Follows directly from Theorems 2.5.3 and 2.5.4. ∎

Given an optimal feasible solution of one of the problems in the primal–dual pair, the above results can be used to characterize the set of all optimal feasible solutions of the other problem.

2.5.3 Economic Interpretation

Consider the primal–dual pair given by (2.40). Assume that the primal problem represents a cost center, such as a government agency, that provides a set of services to its clients, and that the dual represents a consulting firm that can provide the same set of services with various complements of resources. For the former, suppose that the demand for service type i is b_i, $i = 1, \ldots, m$, which can be met with n different types of resources. Let a_{ij} be the amount of service type i provided by one unit of resource j at cost c_j. The problem is to decide how many units of resource j to employ, denoted by x_j, so that all demand is met at minimum cost.

In contrast, the problem faced by the consulting firm is to determine how to price its services to maximize its profit while assuring that the cost to the customer is no greater than he can achieve by employing his own resources to provide the services. To construct a model, let λ_i be the price charged per unit of service type i. The constraints on the prices are $\sum_{i=1}^{m} \lambda_i a_{ij} \leqq c_j$ for all $j = 1, \ldots, n$. Given these prices, the cost center checks whether it can meet its service levels less expensively by contracting with the consulting firm rather than by acquiring or employing its own resources. Because the cost per unit of resource j is c_j, and it can get the same complement of service provided by one unit of resource j form the consulting firm at a cost of $\sum_{i=1}^{m} \lambda_i a_{ij}$, it would not be worthwhile for the cost center to use any amount of resource j if $\sum_{i=1}^{m} \lambda_i a_{ij} < c_j$. In this case, then, we have $x_j = 0$.

Conversely, if the consulting firm associates a positive price per unit of service type i ($\lambda_i > 0$), then the cost center would do best by meeting exactly its service demands. This implies that when $\lambda_i > 0$ the cost center will try to satisfy $\sum_{j=1}^{n} a_{ij} x_j = b_i$. A similar interpretation can be given about the prices that the consulting firm will adopt knowing the amounts of resources that the cost center is using.

These are precisely the complementarity slackness conditions for the primal–dual pair (2.40). When they are satisfied, there is no incentive for the cost center to change the amount of resources it employs or for the consulting firm to change its prices. The minimum cost incurred by the former is exactly the maximum revenue realized by the latter. This results in an economic equilibrium where cost $\left(\sum_{j=1}^{n} c_j x_j\right)$ = revenue $\left(\sum_{i=1}^{m} \lambda_i b_i\right)$.

2.5.4 Sensitivity Analysis

We have just seen that the optimal values of the dual variables in a linear programming problem can be interpreted as prices. In this section this interpretation is explored in further detail.

Suppose in an LP in standard form (2.1) the optimal basis is $\boldsymbol{B}$ with solution $(\boldsymbol{x}_B, \boldsymbol{0})$, where $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$. A solution to the corresponding dual problem is $\boldsymbol{\lambda}^T = \boldsymbol{c}_B\boldsymbol{B}^{-1}$. Now, assuming nondegeneracy, small changes in the vector $\boldsymbol{b}$ will not cause the optimal basis to change. Thus for $\boldsymbol{b} + \Delta\boldsymbol{b}$ the optimal solution is

$$\boldsymbol{x} = (\boldsymbol{x}_B + \Delta\boldsymbol{x}_B, \boldsymbol{0})$$

where $\Delta\boldsymbol{x}_B = \boldsymbol{B}^{-1}\Delta\boldsymbol{b}$. Thus the corresponding increment in the cost function is

$$\Delta f = \boldsymbol{c}_B\Delta\boldsymbol{x}_B = \boldsymbol{\lambda}^T\Delta\boldsymbol{b}$$

This equation shows that $\boldsymbol{\lambda}$ gives the sensitivity of the optimal cost with respect to small changes in the vector $\boldsymbol{b}$. In other words, if a new problem were solved with $\boldsymbol{b}$ changed to $\boldsymbol{b} + \Delta\boldsymbol{b}$, the change in the optimal value of the objective function would be $\boldsymbol{\lambda}^T\Delta\boldsymbol{b}$.

This interpretation says that λ_i directly reflects the change in cost due to a change in the ith component of the vector $\boldsymbol{b}$. Thus λ_i may be viewed equivalently as the *marginal price* of the component b_i, since if b_i is changed to $b_i + \Delta b_i$ the value of the optimal solution changes by $\lambda_i\Delta b_i$. When the constraints $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ are written as $\boldsymbol{A}\boldsymbol{x} \geqq \boldsymbol{b}$ as in (2.40), the dual variables are nonnegative implying that for λ_i positive, a positive change in b_i will produce an increase in the objective function value. In economic terms, it is common to refer to the dual variables as shadow prices.

Example 2.5.1 The shadow prices are associated with constraints but they are often used to evaluate prices or cost coefficients associated with variables of the primal problem. As an example, suppose we have an $\boldsymbol{A}$ matrix which represents the daily operation of an oil refinery and a particular variable x_j representing the purchase of crude oil feedstock, with a cost of \$22.65/barrel ($c_j = 22.65$). The refinery wants to minimize its costs. There is an upper limit on the purchase of this oil of 50,000 barrels/day at this price. This is represented by the constraint

$$x_j + x_s = 50,000$$

where x_s is the associated slack variable. Assume at the optimal x_s has a reduced cost of −\$2.17/barrel: what does this mean?

The shadow price on the constraint is −\$2.17/barrel but this does not mean that we should only pay \$2.17 for another barrel of crude. It means we should be prepared to pay another \$2.17/barrel for an opportunity to purchase extra supplies given that any further purchases would cost \$22.65/barrel; i.e., the objective function will decrease by \$2.17 for every extra barrel we can purchase at the price c_j already in the cost row. This means we should be prepared to bid up to 22.65 + 2.17 = \$25.82/barrel on the spot market for extra supplies of that crude. Note that \$25.82/barrel is the breakeven price, in that we decrease our objective function f if we can purchase a barrel for less than this price, increase f if we purchase for more, and make no change at all to f if we purchase for exactly \$25.82/barrel.

The reduced cost of a variable nonbasic at its lower bound is often referred to as the *opportunity cost* of that variable. If management made the (nonoptimal) decision of increasing that nonbasic from its lower bound, the reduced cost gives the increase in f per unit increase in the variable (for a certain range). This represents the opportunity loss in departing from the optimal solution.

Ranging: For reasons that practitioners understand implicitly, it is often said that postoptimality analysis is the most important part of the LP calculations. The majority of the coefficients that appear in an LP are rarely known with certainty and so have to be estimated from historical or empirical data. Under these circumstances we would like to know the range of variation of these coefficients for which the optimal solution remains optimal; i.e., the basis does not change. Three categories are investigated below: cost coefficients c_j, right-hand-side terms b_i, and matrix coefficients a_{ij}.

Changes in the Cost Row

(a) *Nonbasic variable* The change in the cost coefficient of a nonbasic variable affects the reduced cost of that variable only, and the change is in direct proportion. If δ is a perturbation associated with the original cost coefficient c_q, then at optimality we can write the reduced cost coefficient of nonbasic variable x_q as $\bar{c}_q(\delta) = c_q + \delta - \boldsymbol{\lambda}^T \boldsymbol{a}_q$. In order for the current basis $\boldsymbol{B}$ to remain optimal, we must have $\bar{c}_q(\delta) \geqq 0$. This means

$$\delta \geq \boldsymbol{\lambda}^T \boldsymbol{a}_q - c_q = -\bar{c}_q$$

The reduced cost coefficients of all the other variables are independent of c_q and so will remain nonnegative. If a δ is chosen that violates this inequality, x_q would be identified as the entering variable and we would continue the application of the simplex method until a terminal basis for the modified problem was found.

It is worth mentioning that in most commercial LP codes there is another range given at the same time as well — the range over which x_q can be increased from zero before a change of basis occurs. When $\delta = -\bar{c}_q$, the reduced cost is zero implying that x_q can be increased without affecting the value of the objective function. The maximum value it can take without effecting a change in basis is given by $\min_i\{\bar{b}_i/\alpha_{iq} : \alpha_{iq} > 0\}$ which is the minimum ratio test in Step 3 of the simplex method.

(b) *Basic variable* A change in the cost coefficient of a basic variable may affect the reduced cost of all the nonbasic variables. Let $\boldsymbol{e}_i$ be the ith unit vector of length m and suppose we increment the cost coefficient of the ith basic variable by δ; i.e., $\boldsymbol{c}_B \leftarrow \boldsymbol{c}_B + \delta \boldsymbol{e}_i^T$. This gives $\boldsymbol{\lambda}^T(\delta) = (\boldsymbol{c}_B + \delta \boldsymbol{e}_i^T)\boldsymbol{B}^{-1}$ so the dual vector is an affine function of δ. The reduced cost of the qth nonbasic variable is now

$$\begin{aligned} \bar{c}_q(\delta) &= c_q - (\boldsymbol{c}_B + \delta \boldsymbol{e}_i^T)\boldsymbol{B}^{-1}\boldsymbol{a}_q \\ &= c_q - \boldsymbol{c}_B \boldsymbol{B}^{-1}\boldsymbol{a}_q - \delta \boldsymbol{e}_i^T \boldsymbol{B}^{-1}\boldsymbol{a}_q \\ &= \bar{c}_q - \delta \alpha_{iq} \end{aligned}$$

where $\alpha_{iq} = (\boldsymbol{B}^{-1}\boldsymbol{a}_q)_i$ is the ith component of the updated column $\boldsymbol{a}_q$. This value is found for the nonbasic variable x_q by solving $\boldsymbol{B}^T\boldsymbol{y} = \boldsymbol{e}_i$ for $\boldsymbol{y}$, then computing $\alpha_{iq} = \boldsymbol{y}^T\boldsymbol{a}_q$. (Obviously, if $\alpha_{iq} = 0$ for any x_q, the reduced cost does not change.)

For a solution to remain optimal, we must have $\bar{c}_q(\delta) \geqq 0$, or

$$\bar{c}_q - \delta\alpha_{iq} \geqq 0 \ \forall q \tag{2.46}$$

where $\bar{c}_q$ is the reduced cost at the current optimum. This constraint produces bounds on δ. For a basic variable, the *range* over which c_i can vary and the current solution remain optimal is given by $c_i + \delta$, where

$$\max_q \left\{ \frac{\bar{c}_q}{\alpha_{iq}} \;\middle|\; \alpha_{iq} < 0 \right\} \leqq \delta \leqq \min_q \left\{ \frac{\bar{c}_q}{\alpha_{iq}} \;\middle|\; \alpha_{iq} > 0 \right\}$$

since this is the range for which (2.46) is satisfied. If there is no $\alpha_{iq} > 0$, then $\delta < \infty$ likewise if there is no $\alpha_{ij} < 0$, then $\delta > -\infty$.

Example 2.5.2 Suppose we have an optimal solution to an LP given in tableau form with attached variables

$$\begin{array}{lrcl} \min\limits_{\boldsymbol{x}\geqq\boldsymbol{0}} & -f = & -31.5 - 3.5x_4 - 0.1x_3 - 0.25x_5 & \\ \text{subject to} & x_1 = & 3.2 - 1.0x_4 - 0.5x_3 - 0.6x_5 & \\ & x_2 = & 1.5 + 0.5x_4 + 1.0x_3 - 1.0x_5 & \\ & x_6 = & 5.6 - 2.0x_4 - 0.5x_3 - 1.0x_5 & \end{array} \tag{2.47}$$

If the cost coefficient of x_2 becomes $c_2 + \delta$, the reduced costs of the nonbasic variables become:

$$\begin{aligned} x_4 &: \ \bar{c}_4(\delta) = 3.5 - \delta(-0.5) \\ x_3 &: \ \bar{c}_3(\delta) = 0.1 - \delta(-1.0) \\ x_5 &: \ \bar{c}_5(\delta) = 0.25 - \delta(+1.0) \end{aligned}$$

Note that $x_{B_i} = \bar{b}_i - \sum_{j\in\{3,4,5\}} \alpha_{ij}x_j$ for $i = 1, 2, 6$, so α_{ij} is the negative of the number appearing in equations (2.47). The range that δ can take is given by

$$\max\left\{ \frac{3.5}{-0.5}, \frac{0.1}{-1.0} \right\} \leqq \delta \leqq \min\left\{ \frac{0.25}{1.0} \right\}$$

$$-0.1 \leqq \delta \leqq 0.25$$

When δ assumes one of the limits of its range, a reduced cost becomes zero. In this example, for $\delta = -0.1$ the reduced cost of x_3 is zero, so that if the cost coefficient of x_2 decreases by 0.1 or more it becomes advantageous for x_3 to become active. The minimum ratio test: $\min\{3.2/0.5,\ 5.6/0.5\}$ then indicates that 6.4 units of x_3 can be produced before x_1 becomes zero and a change of basis is required.

Changes in the Right-Hand-Side Vector

We wish to investigate the effect of a change $b_i \leftarrow b_i + \delta$ for some $1 \leqq i \leqq m$. It is usual to consider the case where b_i is the right-hand side of an inequality constraint, which therefore has a slack variable associated with it. The goal is to determine the range over which the current solution remains optimal. If the constraint is an equality, it can be analyzed by regarding its associated artificial variable as a positive slack (which must be nonbasic for a feasible solution).

(a) *Basic slack variable:* If the slack variable associated with the ith constraint is basic the constraint is not binding at the optimum. The analysis is simple: the value of the slack gives the range over which the right-hand side b_i can be reduced (increased for $\geqq$ constraint). The solution remains feasible and optimal for the range $b_i + \delta$, where

$$\begin{aligned} \widehat{x}_s \leqq \delta < \infty \qquad & \text{for } \leqq \text{ type constraint} \\ -\infty < \delta \leqq \widehat{x}_s \qquad & \text{for } \geqq \text{ type constraint} \end{aligned}$$

where $\widehat{x}_s$ is the value of the associated slack variable.

(b) *Nonbasic slack variable:* If a slack variable is nonbasic at zero, then the original inequality constraint is binding at the optimum. At first sight it would seem that because the constraint is binding, there is no possibility of changing the right-hand side term, particularly in decreasing the value of b_i (for $\leqq$ type constraints). It turns out that by changing the vector $\boldsymbol{b}$ we also change $\boldsymbol{x}_B$ $(= \boldsymbol{B}^{-1}\boldsymbol{b})$ so there is a range over which $\boldsymbol{x}_B$ remains nonnegative. For the associated values we still retain an optimal feasible solution in the sense that the basis does not change. (Note that both $\boldsymbol{x}_B$ and $f = \boldsymbol{c}_B\boldsymbol{x}_B$ change value.)

Consider the constraint

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + x_s = b_k \tag{2.48}$$

where x_s is the slack variable. If the right-hand side becomes $b_k + \delta$, rearranging (2.48) gives

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + (x_s - \delta) = b_k \tag{2.49}$$

so that $(x_s - \delta)$ replaces x_s. Thus if x_s is nonbasic at zero in the final tableau, we have the expression

$$\boldsymbol{x}_B = \bar{\boldsymbol{b}} - \boldsymbol{\alpha}_s(-\delta)$$

where $\boldsymbol{\alpha}_s$ is the updated column in the tableau corresponding to x_s. Because $\boldsymbol{x}_B$ must remain nonnegative, we have $\bar{\boldsymbol{b}} + \delta\boldsymbol{\alpha}_s \geqq \boldsymbol{0}$ which is used to solve for the range over which δ can vary.

$$\max_i \left\{ \frac{\bar{b}_i}{-\alpha_{is}} \;\middle|\; \alpha_{is} > 0 \right\} \leqq \delta \leqq \min_i \left\{ \frac{\bar{b}_i}{-\alpha_{is}} \;\middle|\; \alpha_{is} < 0 \right\}$$

If there is no $\alpha_{is} > 0$, then $\delta > -\infty$; if there is no $\alpha_{is} < 0$, then $\delta < \infty$.

For $\geqq$ type constraints, δ changes sign. This follows because we can analyze $\sum_{j=1}^{n} a_{ij}x_j \geqq b_i$ in the form $-\sum_{j=1}^{n} a_{ij}x_j \leqq -b_i$, so that $-(x_s + \delta)$ replaces $(x_s - \delta)$ in eq. (2.49). Another way of seeing this is to consider the change to the right-hand side in the form

$$\boldsymbol{b}(\delta) = \boldsymbol{b} + \delta \boldsymbol{e}_k$$

Thus the new value of $\boldsymbol{x}_B$ is given by

$$\begin{aligned} \boldsymbol{x}_B(\delta) &= \boldsymbol{B}^{-1}\boldsymbol{b}(\delta) = \boldsymbol{B}^{-1}\boldsymbol{b} + \delta B^{-1}\boldsymbol{e}_k \\ &= \bar{\boldsymbol{b}} + \delta \boldsymbol{B}^{-1}\boldsymbol{e}_k \end{aligned}$$

But as

$$\boldsymbol{\alpha}_s = \boldsymbol{B}^{-1}\boldsymbol{e}_k \quad \text{for a } \leqq \text{ type constraint}$$

and as

$$\boldsymbol{\alpha}_s = -\boldsymbol{B}^{-1}\boldsymbol{e}_k \quad \text{for a } \geqq \text{ type constraint}$$

since the column corresponding to the slack variable is $+\boldsymbol{e}_k$ for a $\leqq$ constraint and $-\boldsymbol{e}_k$ for a $\geqq$ constraint. Thus we have

$$\begin{aligned} \bar{\boldsymbol{b}} - \boldsymbol{\alpha}_s(-\delta) \geqq \boldsymbol{0} \quad &\text{for a } \leqq \text{ type constraint, and} \\ \bar{\boldsymbol{b}} - \boldsymbol{\alpha}_s(+\delta) \geqq \boldsymbol{0} \quad &\text{for a } \geqq \text{ type constraint} \end{aligned}$$

Example 2.5.3 Consider Example 2.5.2 again and suppose x_4 represents a slack variable for a particular constraint i ($\leqq$ type). If the coefficient b_i is varied by an amount δ, we have

$$\begin{aligned} x_1(\delta) &= 3.2 - 1.0(-\delta) \quad \text{that is, } \bar{\boldsymbol{b}} = (3.2, 1.5, 5.6)^T \\ x_2(\delta) &= 1.5 + 0.5(-\delta) \qquad\qquad \boldsymbol{\alpha}_s = (1.0, -0.5, 2.0)^T \\ x_6(\delta) &= 5.6 - 2.0(-\delta) \end{aligned}$$

Thus

$$\begin{aligned} x_1(\delta) \geqq 0 \text{ for } 3.2 - 1.0(-\delta) \geqq 0, \text{ that is, } \delta &\geqq \frac{3.2}{-1.0} \\ x_2(\delta) \geqq 0 \text{ for } 1.5 + 0.5(-\delta) \geqq 0, \text{ that is, } \delta &\leqq \frac{1.5}{0.5} \\ x_6(\delta) \geqq 0 \text{ for } 5.6 - 2.0(-\delta) \geqq 0, \text{ that is, } \delta &\geqq \frac{5.6}{-2.0} \end{aligned}$$

Therefore, δ can vary in the range

$$\begin{aligned} \max\left\{\frac{3.2}{-1.0}, \frac{5.6}{-2.0}\right\} \leqq \delta \leqq \min\left\{\frac{1.5}{0.5}\right\} \\ -2.8 \leqq \delta \leqq 3.0 \end{aligned}$$

Changes in Matrix Coefficients

The technological coefficients a_{ij} are usually known with much more certainty than the cost row or right-hand-side vector, since they customarily represent some physical interaction between variables and are not subject to the same market fluctuations as costs and demands. We shall consider changes to the coefficients of nonbasic variables only; changes to coefficients of basic variables alters the basis matrix $\boldsymbol{B}$ and are rather complicated to analyze (see [M18]).

Consider the jth nonbasic variable with corresponding column $\boldsymbol{a}_j$. If the ith element of $\boldsymbol{a}_j$ is changed by an amount δ, this affects the reduced cost $\bar{c}_j$ as follows.

If

$$\boldsymbol{a}_j(\delta) = \boldsymbol{a}_j + \delta \boldsymbol{e}_i$$

then

$$\begin{aligned} \bar{c}_j(\delta) &= c_j - \boldsymbol{\lambda}^T(\boldsymbol{a}_j + \delta \boldsymbol{e}_j) \\ &= \bar{c}_j - \delta \boldsymbol{\lambda}^T \boldsymbol{e}_i \\ &= \bar{c}_j - \delta \lambda_i \end{aligned}$$

where $\boldsymbol{\lambda}^T (= \boldsymbol{c}_B \boldsymbol{B}^{-1})$ is the dual vector. Thus the solution remains optimal as long as $\bar{c}_j(\delta) \geqq 0$. The corresponding range for δ is

$$\delta \leqq \frac{\bar{c}_j}{\lambda_i} \quad \text{for } \lambda_i > 0$$

$$\delta \geqq \frac{\bar{c}_j}{\lambda_i} \quad \text{for } \lambda_i < 0$$

2.5.5 Dual Simplex Method

The guiding principle of the simplex method is to maintain primal feasibility ($x_{B_i} = \bar{b}_i \geqq 0$, $i = 1, \ldots, m$) and complementary slackness ($x_j \bar{c}_j = 0$, for all j) at each iteration while working toward nonnegative reduced costs ($\bar{c}_j \geqq 0$, for all j). This last condition is known as dual feasibility because it is equivalent to satisfying the dual constraints $\boldsymbol{\lambda}^T \boldsymbol{A} \leqq \boldsymbol{c}$, where $\boldsymbol{\lambda}^T = \boldsymbol{c}_B \boldsymbol{B}^{-1}$. Any basis that satisfies primal feasibility and dual feasibility automatically satisfies the complementary slackness condition and is hence optimal. In this section, we present the dual simplex algorithm which maintains dual feasibility and complementarity at each iteration while striving for primal feasibility.

We are interested in such an algorithm because often we have available a basic solution to an LP that is not feasible but which prices out optimally; that is, the corresponding multipliers are feasible for the dual problem. In the simplex tableau this situation

corresponds to having no negative reduced cost coefficients but an infeasible basic solution. Such a situation may arise, for example, if a solution to a certain LP is calculated and then a new problem is constructed by changing the vector $\boldsymbol{b}$. Here, a basic feasible solution to the dual is available and hence it is desirable to pivot in such a way as to optimize the dual.

Rather than constructing a tableau for the dual problem (which, if the primal is in standard form, involves m free variables and n nonnegative slack variables), it is more efficient to work on the dual from the primal tableau. The complete technique based on this idea is the dual simplex method.

Given the linear programming problem in standard form: $\min\{\boldsymbol{cx} : \boldsymbol{Ax} = \boldsymbol{b},\ \boldsymbol{x} \geqq \boldsymbol{0}\}$, suppose a basis $\boldsymbol{B}$ is known such that $\boldsymbol{\lambda}^T = \boldsymbol{c}_B\boldsymbol{B}^{-1}$ is feasible for the dual. In this case we say that the corresponding basic solution to the primal, $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$, is *dual feasible*. If $\boldsymbol{x}_B \geqq \boldsymbol{0}$ then this solution is also primal feasible and hence optimal.

The given vector $\boldsymbol{\lambda}$ is feasible for the dual and thus satisfies $\boldsymbol{\lambda}^T\boldsymbol{a}_j \leq c_j$, for $j = 1, \ldots, n$. Indeed, assuming as usual that the basis is the first m columns of $\boldsymbol{A}$, we have m equalities

$$\boldsymbol{\lambda}^T\boldsymbol{a}_j = c_j, \quad j = 1, \ldots, m$$

and (barring degeneracy in the dual) $n - m$ inequalities

$$\boldsymbol{\lambda}^T\boldsymbol{a}_j < c_j, \quad j = m+1, \ldots, n$$

To develop one cycle of the dual simplex method, we find a new vector $\widehat{\boldsymbol{\lambda}}$ such that one of the equalities becomes an inequality and one of the inequalities becomes an equality, while at the same time increasing the value of the dual objective function. The m equalities in the new solution then determine a new basis.

Denote the ith row of $\boldsymbol{B}^{-1}$ by $\boldsymbol{\beta}_i$. Then for

$$\widehat{\boldsymbol{\lambda}}^T = \boldsymbol{\lambda}^T - \theta\boldsymbol{\beta}_i$$

we have $\widehat{\boldsymbol{\lambda}}^T\boldsymbol{a}_j = \boldsymbol{\lambda}^T\boldsymbol{a}_j - \theta\boldsymbol{\beta}_i\boldsymbol{a}_j$. Noting that $\boldsymbol{\beta}_i\boldsymbol{a}_j = \alpha_{ij}$, the ijth element of the tableau, we have

$$\widehat{\boldsymbol{\lambda}}^T\boldsymbol{a}_j = c_j, \quad j = 1, \ldots, m;\ j \neq i \tag{2.50a}$$

$$\widehat{\boldsymbol{\lambda}}^T\boldsymbol{a}_j = c_i - \theta \tag{2.50b}$$

$$\widehat{\boldsymbol{\lambda}}^T\boldsymbol{a}_j = \boldsymbol{\lambda}^T\boldsymbol{a}_j - \theta\alpha_{ij}, \quad j = m+1, \ldots, n \tag{2.50c}$$

Also,

$$\widehat{\boldsymbol{\lambda}}^T\boldsymbol{b} = \boldsymbol{\lambda}^T\boldsymbol{b} - \theta\boldsymbol{x}_{B_i} \tag{2.51}$$

These last equations lead directly to the algorithm.

Step 1 Given a dual feasible basic solution $\boldsymbol{x}_B$, if $\boldsymbol{x}_B \geqq \mathbf{0}$ the solution is optimal. If $\boldsymbol{x}_B$ contains one or more negative components, select an index i such that $x_{B_i} < 0$.

Step 2 If all $\alpha_{ij} \geqq 0$, $j = 1, \ldots, n$, then the dual has no maximum (this follows from (2.50) because $\boldsymbol{\lambda}$ is feasible for all $\theta > 0$). If $\alpha_{ij} < 0$ for some j, let

$$\theta = \frac{\bar{c}_p}{-\alpha_{ip}} = \min_j \left\{ \frac{\bar{c}_j}{-\alpha_{ij}} \;\middle|\; \alpha_{ij} < 0 \right\} \tag{2.52}$$

where p is the index that corresponds to the minimum ratio.

Step 3 Form a new basis $\boldsymbol{B}$ by replacing $\boldsymbol{a}_i$ with $\boldsymbol{a}_p$. Using this basis determine the corresponding basic dual feasible solution $\boldsymbol{x}_B$ and return to Step 1.

At Step 1 we choose a leaving variable, usually the most negative; at Step 2 we choose an entering variable based on a ratio test whose primal counterpart is (2.23). The proof that the algorithm converges to the optimal solution is similar in detail to the proof for the primal simplex method. The essential observations are: (a) from the choice of p in (2.52) and from (2.50) the new solution will again be dual feasible; (b) by (2.51) and the choice $x_{B_i} < 0$, the value of the dual objective will increase; (c) the procedure cannot terminate at a nonoptimal point; and (d) since there are only a finite number of bases, the optimum must be achieved in a finite number of steps when a mechanism for taking degeneracy into account is included.

Example 2.5.4 A form of problem arising frequently is that of minimizing a positive combination of nonnegative variables subject to a series of "greater than" type inequalities having positive coefficients. Such problems, as given below, are natural candidates for application of the dual simplex procedure.

$$\begin{aligned} \min\ f \;=\quad & 3x_1 + 4x_2 + 5x_3 \\ \text{subject to}\quad & x_1 + 2x_2 + 3x_3 \geqq 5 \\ & 2x_1 + 2x_2 + x_3 \geqq 6 \\ & x_1,\ x_2,\ x_3 \geqq 0 \end{aligned}$$

By introducing surplus variables x_4 and x_5, and by changing the sign of the inequalities we obtain the initial tableau:

	x_1	x_2	x_3	x_4	x_5	$\bar{\boldsymbol{b}}$
x_4	-1	-2	-3	1	0	-5
x_5	$\textcircled{-2}$	-2	-1	0	1	-6
$-f$	3	4	5	0	0	0

Initial tableau

The basis corresponds to a dual feasible solution since all of the $\bar{c}_j$'s are nonnegative. We select any $x_{B_i} < 0$, say $x_5 = -6$, to remove from the set of basic variables. To find the appropriate pivot element in the second row we compute the ratios $-\bar{c}_j/\alpha_{2j}$ and select the minimum positive value: $\min\{3/2,\ 4/2,\ 5/1\} = 1.5$ corresponding to x_1. This yields the indicated pivot. Continuing, the remaining tableaus are

	x_1	x_2	x_3	x_4	x_5	$\bar{\boldsymbol{b}}$
x_4	0	(−1)	−5/2	1	−1/2	−2
x_1	1	1	1/2	0	−1/2	3
$-f$	0	1	7/2	0	3/2	−9

Second tableau

	x_1	x_2	x_3	x_4	x_5	$\bar{\boldsymbol{b}}$
x_2	0	1	5/2	−1	1/2	2
x_1	1	0	−2	1	−1	1
$-f$	0	0	1	1	1	−11

Final tableau

The third tableau yields a feasible solution to the primal which must be optimal. Thus the solution is $x_1 = 1$, $x_2 = 2$, $x_3 = 0$.

REFERENCES

[B15] M.S. Bazaraa, J.J. Jarvis and H.D. Sherali, *Linear Programming and Network Flows*, Second Edition, John Wiley & Sons, New York (1990).

[B22] D. Bertsimas and J.N.Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA (1997).

[B25] R.E. Bixby, "Progress in Linear Programming," *ORSA Journal on Computing*, Vol. 6, No. 1, pp. 15–22 (1994).

[C4] I.C. Chow, C.L. Monma and D.F. Shanno, "Further Development of a Primal-Dual Interior Point Method," *ORSA Journal on Computing*, Vol. 2, No. 4, pp. 304–311 (1990).

[G2] S.I. Gass, *Linear Programming: Methods and Applications*, Fifth Edition, McGraw Hill, New York (1985).

[H3] P.M.J. Harris, "Pivot Selection Methods for the Devex LP Code," *Mathematical Programming*, Vol. 5, pp. 1–28 (1973).

[H9] J.N. Hooker, "Karmarkar's Linear Programming Algorithm," *Interfaces*, Vol. 16, No. 4, pp. 75–90 (1986).

[K2] N. Karmarkar, "A New Polynomial-Time Algorithm for Linear Programming," *Combinatorica*, Vol. 4, pp. 373–395 (1984).

[L8] I.J. Lustig, R.E. Marsten and D.F Shanno, "Computational Experience with a Primal-Dual Interior Point Method," *Linear Algebra and Its Applications*, Vol. 152, pp. 191–222 (1991).

[M7] R.E. Marsten, M.J. Saltzman, D.F. Shanno, G.S. Pierce and J.F. Ballintijn, "Implementation of a Dual Affine Interior Point Algorithm for Linear Programming," *ORSA Journal on Computing*, Vol. 1, No. 4, pp. 287–297 (1989).

[M8] R.E. Marsten, R. Subramanian, M. Saltzman, I. Lustig and D. Shanno, "Interior Point Methods for Linear Programming: Just Call Newton, Lagrange, and Fiacco and McCormick," *Interfaces*, Vol. 20, No. 4, pp. 105–116 (1990).

[M9] T.H. Matheiss and D.S. Rubin, "A Survey and Comparison of Methods for Finding all Vertices of a Convex Polyhedral Sets," *Operations Research*, Vol. 5, No. 2, pp. 167–185 (1980).

[M12] P.G. McKeown, "A Vertex Ranking Procedure for Solving the Linear Fixed Charge Problem," *Operations Research*, Vol. 23, No. 6, pp. 1183–1191 (1975).

[M17] K.G. Murty, "Adjacency on Convex Polyhedra," *SIAM Review*, Vol. 13, No. 3, pp. 377–386 (1971).

[M18] K.G. Murty, *Linear Programming*, John Wiley & Sons, New York (1983).

[S17] U.H. Suhl and L.M. Suhl, "Computing Sparse LU Factorizations for Large-Scale Linear Programming Bases," *ORSA Journal on Computing*, Vol. 2, No. 4, pp. 325–335 (1994).

[V1] R.J. Vanderbei, *Linear Programming: Foundations and Extensions*, Kluwer Academic Publishers, Boston (1996).