2. LINEAR EQUATIONS AND INEQUALITIES

Overview

As we have observed, the constraints of a meaningful linear program must include at least one linear inequality, but otherwise they may be composed of linear equations, linear inequalities, or some of each. In this chapter, we touch on two important topics within the very large subject of linear equations and linear inequalities.

The first of these topics concerns *equivalent forms of the linear programming problem.* To appreciate this subject, it is a good idea to think about what we mean by the noun "form". *Webster's New Collegiate Dictionary* defines form as "the shape and structure of something as distinguished from its material." In the present context, this has much to do with the types of mathematical structures (sets, functions, equations, inequalities, etc.) used to represent an object, such as a linear programming problem. It is equally important to have a sense of what it means for two forms of an optimization problem (or two optimization problems) to be "equivalent." This concept is a bit more subtle. Essentially, it means that there is a one-to-one correspondence between the optimal solutions of one problem and those of the other. Using the equivalence of optimization problems, one can sometimes achieve remarkable computational efficiencies. In this chapter we illustrate another use of equivalence by converting two special nonlinear optimization problems to linear programs (see Example 2.1 on page 35 and Example 2.2 on page 38).

The second important topic introduced in this chapter is about *properties of polyhedral convex sets*, by which we mean solution sets of linear inequality systems. In discussing the geometry of polyhedral sets it is customary to use words such as *hyperplane, halfspace, vertex, and edge.* In this chapter we will cover these geometric concepts and relate them to algebraic structures.

The geometric study of polyhedral convex sets goes back to classical antiquity. Much of what is now known about this subject makes use of algebraic methods and was found in more "recent" times, which is to say the 19th and 20th centuries.

In learning this material we gain valuable knowledge and useful vocabulary for practical and theoretical work alike.

R.W. Cottle and M.N. Thapa, *Linear and Nonlinear Optimization*,
International Series in Operations Research & Management Science 253,
DOI 10.1007/978-1-4939-7055-1_2

2.1 Equivalent forms of the LP

The material covered in this section includes: the conversion of a linear inequality to an equivalent linear equation plus a nonnegative variable; the equivalence of a linear equation to a pair of linear inequalities; the transformation of variables that are not required to be nonnegative to variables that are so restricted; and, finally, the conversion of a maximization problem to an equivalent minimization problem.

In Chapter 1, we introduced the term "standard form" for a linear program: A linear programming problem will be said to be in *standard form* if its constraints are all equations and its variables are all required to be nonnegative. Such a problem has only constraints of the form

$$\begin{aligned} \sum_{j=1}^{n} a_{ij}x_j &= b_i, \qquad i = 1, \dots, m \\ x_j &\geq 0, \qquad j = 1, \dots, n. \end{aligned}$$

More tersely, the row-oriented presentation of the standard form constraints written above has the matrix form

$$\begin{aligned} Ax &= b \\ x &\geq 0. \end{aligned}$$

Here A denotes an $m \times n$ matrix, and b is an m-vector.

The transportation problem discussed in Example 1.1 was presented in standard form. Nevertheless, as we have seen in the other examples discussed in Chapter 1, some linear programs are not initially in standard form. Granting for the moment that there is something advantageous about having a linear program in standard form, we look at the matter of converting the constraints of a linear program to standard form when they don't happen to be that way initially. In so doing, we need to be sure that we do not radically alter the salient properties of the problem. This statement alludes to the notion of *equivalence* of optimization problems. We will have occasion to use the concept of equivalence later in this and subsequent chapters.

But what is to be done with linear programs having constraints that are not in standard form? For instance, the diet problem (Example 1.2), the product-mix problem (Example 1.3) and the blending problem (page 13) all have linear inequalities among the *functional constraints*, i.e., those which are not just nonnegativity conditions or *bound constraints* as they are called.

Conversion to standard form

A single linear inequality of the form

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i$$

can be converted to a linear equation by adding a nonnegative *slack variable.* Thus, the inequality above is equivalent to

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i, \quad x_{n+i} \geq 0.$$

In this case, x_{n+i} is the added slack variable.

Here is a simple illustration of this process. Suppose we have the linear inequality

$$3x_1 - 5x_2 \leq 15. \tag{2.1}$$

At the moment, we are not imposing a nonnegativity constraint on the variables x_1 and x_2. This linear inequality can be converted into a linear equation by adding a nonnegative variable, x_3 to the left-hand side so as to obtain

$$3x_1 - 5x_2 + x_3 = 15, \quad x_3 \geq 0. \tag{2.2}$$

Figure 2.1 depicts how the sign of the slack variable x_3 behaves with respect to the values of x_1 and x_2 in (2.2). The shaded region corresponds to the values of x_1 and x_2 satisfying the given linear inequality.

A system of $\leq$ linear inequalities can be converted to a system of linear equations by adding a separate nonnegative slack variable in each linear inequality. Thus

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i, \quad i = 1, \ldots, m$$

becomes

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i, \quad x_{n+i} \geq 0, \quad i = 1, \ldots, m.$$

To see the importance of using a *separate* slack variable for each individual linear inequality in the system, consider the following system

$$\begin{aligned} 3x_1 - 5x_2 &\leq 15 \\ 2x_1 + 4x_2 &\leq 16. \end{aligned} \tag{2.3}$$

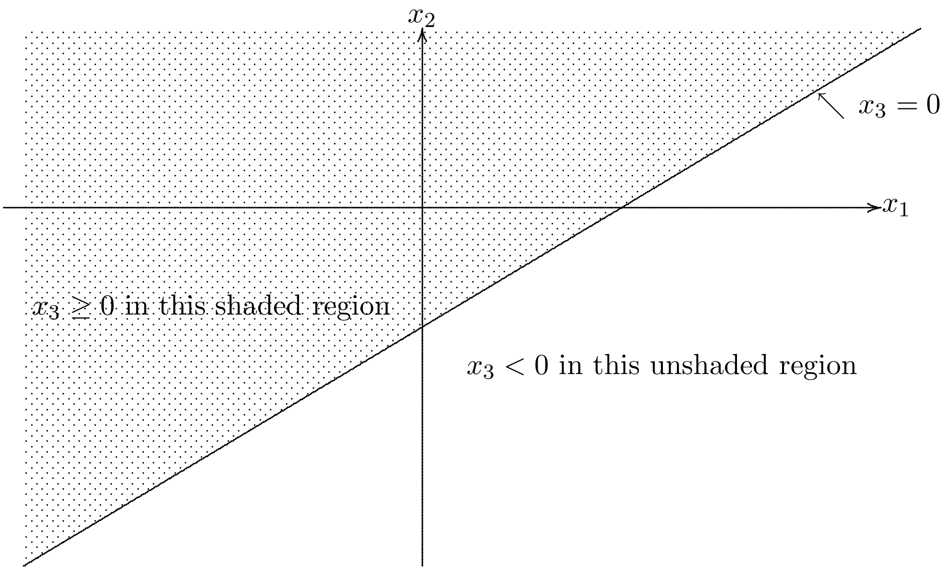

Figure 2.1: Solutions of linear inequality (2.1).

Giving each inequality a slack variable, we obtain the system

$$\begin{array}{lll} 3x_1 - 5x_2 + x_3 = 15, & & x_3 \geq 0 \\ 2x_1 + 4x_2 + x_4 = 16, & & x_4 \geq 0. \end{array} \tag{2.4}$$

Solutions of these linear inequalities are shown in the darkest region of Figure 2.2 below.

In Figure 2.2, it is plain to see that there are solutions of one inequality which are not solutions of the other. This, in turn, has a direct relationship with the *signs* of x_3 and x_4. In short, it would be incorrect to use just one slack variable for the two linear inequalities.

Suppose we now introduce nonnegativity constraints on the variables x_1 and x_2 in the inequality system (2.3). The system could then be written as

$$\begin{array}{c} 3x_1 - 5x_2 + x_3 = 15 \\ 2x_1 + 4x_2 + x_4 = 16 \\ x_j \geq 0, \quad j = 1, 2, 3, 4. \end{array} \tag{2.5}$$

The feasible solutions of this system are indicated by the most darkly shaded region in Figure 2.3.

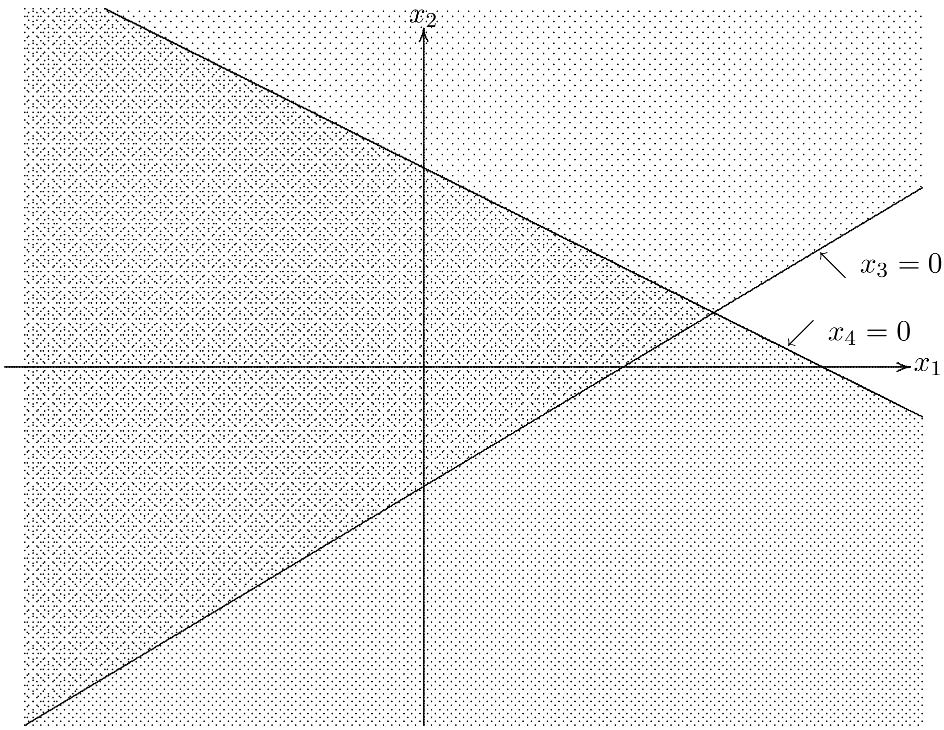

Figure 2.2: Solutions of (2.3).

The same conversion in matrix notation is as follows. Constraints

$$Ax \leq b$$

can be written as

$$Ax + Is = b, \quad s \geq 0$$

where, componentwise, $s_i = x_{n+i}$. Analogously, the constraints

$$Ax \geq b$$

can be written as

$$Ax - Is = b, \quad s \geq 0$$

where, componentwise, $s_i = x_{n+i}$. The components of the vector s are sometimes called *surplus variables*.

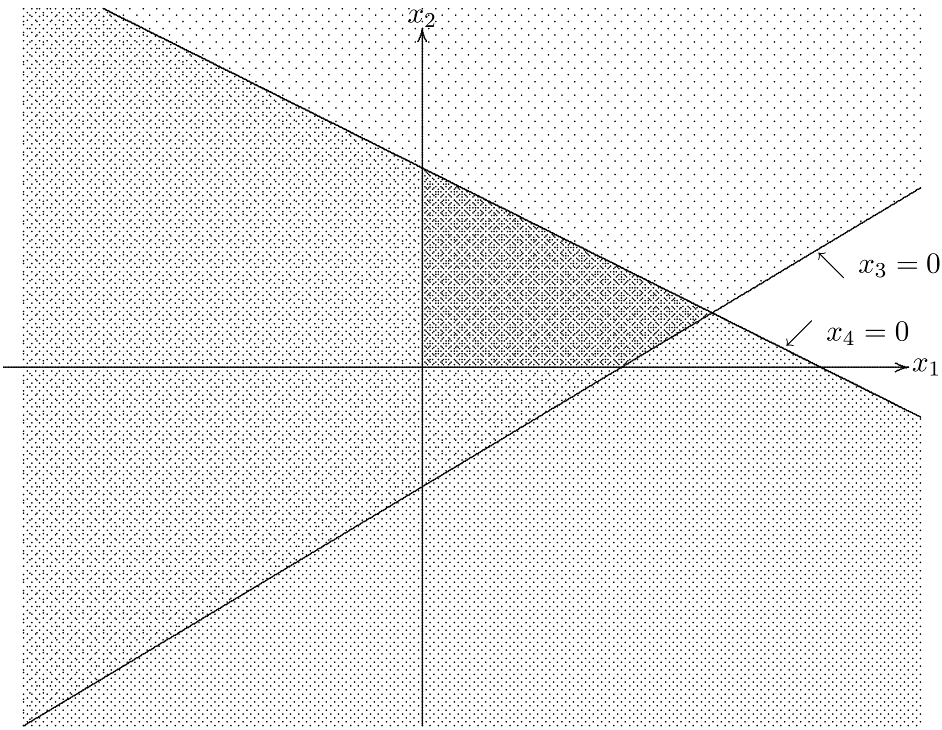

Figure 2.3: Nonnegative solutions of (2.3).

Linear programs with free variables

As you will recall, a linear programming problem is an optimization problem in which a linear function is to be minimized (or maximized) subject to a system of linear constraints (equations or inequalities) on its variables. We insist that the system contain at least one linear inequality, but we impose no further conditions on the number of equations or inequalities. In particular, the variables of a linear program need not be nonnegative. Some linear programming problems have variables that are *unrestricted in sign*. Such variables are said to be *free*. Why do we care whether variables are free or not? This has to do with the fact that the Simplex Algorithm (which we shall take up in Chapter 3) is designed to solve linear programs in standard form.

In the following example, we encounter a classic optimization problem arising in statistics that does not appear to be a linear program. The problem can, however, be converted to a linear program, albeit one that is definitely not in standard form. In particular, it will have linear inequality constraints *and* free variables. Later, we show how to bring this linear program into standard form.

Example 2.1: THE CHEBYSHEV PROBLEM[1]**.** Suppose we have a hypothesis that a particular variable is a linear function of certain variables (that is, a linear model). More specifically, suppose we believe that there are numbers $x_1, x_2, \ldots, x_n$ such that when the inputs[2] $a_1, a_2, \ldots, a_n$ are supplied, the output

$$b = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

will be observed. We would like to know what the numbers $x_1, x_2, \ldots, x_n$ are, so we run an experiment: We select values for the inputs and observe the output. In fact, we do this m times. On the i-th trial, we use inputs $a_{i1}, a_{i2}, \ldots, a_{in}$ and observe the output b_i. The question arises: Does there exist a single set of values for the numbers $x_1, x_2, \ldots, x_n$ such that all the equations

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i, \quad i = 1, \ldots, m$$

hold? If not, can we choose a set of numbers $x_1, x_2, \ldots, x_n$ such that the largest deviation is minimized? By "largest deviation" we mean

$$z := \max\left\{ \left|\sum_{j=1}^{n} a_{1j}x_j - b_1\right|, \ \left|\sum_{j=1}^{n} a_{2j}x_j - b_2\right|, \ \ldots, \ \left|\sum_{j=1}^{n} a_{mj}x_j - b_m\right| \right\}.$$

We want to choose the "weights" $x_1, x_2, \ldots, x_n$ so as to minimize z.

For the sake of exposition, let us assume $m = 2$ and $n = 3$. Here is how such a Chebyshev problem can be turned into a linear program. First note that by definition

$$z \geq |a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1| \quad \text{and} \quad z \geq |a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2|.$$

Since each of these 2 absolute values is nonnegative, we have $z \geq 0$, although we shall not impose this explicitly as a constraint. Furthermore, since for any real number t, $|t| \geq \pm t$, we have

$$|a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1| \geq \pm(a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1)$$

$$|a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2| \geq \pm(a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2).$$

[1]For a treatment of this linear programming model and many others, see [77].

[2]Inputs are also called the independent variables and the output is called the dependent variable.

Hence we obtain the linear program

$$\begin{array}{ll} \text{minimize} & z \\ \text{subject to} & z - (a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1) \geq 0 \\ & z + (a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1) \geq 0 \\ & z - (a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2) \geq 0 \\ & z + (a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2) \geq 0. \end{array}$$

The linear program in general is to minimize z subject to this system of $2m$ linear inequalities in which all the variables—including z—are free. Nevertheless, by our earlier discussion, we know that for any solution of these inequalities, we must have $z \geq 0$.

Conversion of free variables to differences of nonnegative variables

A free variable can always be replaced by the *difference* of two nonnegative variables. Thus

$$\theta = \theta' - \theta'', \quad \theta' \geq 0, \ \theta'' \geq 0.$$

Notice that if $\theta = \theta' - \theta''$, it is not automatically true that the product of θ' and θ'' is zero. (For example, $-1 = 2 - 3$.) Nevertheless, θ' and θ'' can be chosen so that $\theta'\theta'' = 0$. When this is done, we have

$$\theta' = \theta^+ \quad \text{and} \quad \theta'' = \theta^-,$$

where, by definition,

$$\theta^+ = \max\{0, \theta\} \quad \text{and} \quad \theta^- = -\min\{0, \theta\}.$$

To illustrate, suppose $\theta = 5$. Then $\theta^+ = 5$ and $\theta^- = 0$. On the other hand, if $\theta = -5$, then $\theta^+ = 0$ and $\theta^- = 5$. If $\theta = 0$, then $\theta^+ = \theta^- = 0$. In each case, $\theta^+\theta^- = 0$.

Substituting for free variables

If x_j is a free variable and we use the substitution $x_j = x'_j - x''_j$, then, in the objective function, the term c_jx_j becomes $c_jx'_j - c_jx''_j$, and in a constraint, the term $a_{ij}x_j$ becomes $a_{ij}x'_j - a_{ij}x''_j$.

We left Example 2.1 in the form of an inequality-constrained LP with free variables. Using the techniques described above, you can convert this linear program to one in standard form.

Important notational conventions

We often need to single out a row or column of a matrix. Accordingly, it is useful to have a clear and consistent way to do this. There are several such systems in use, but the one described in the following is preferred in this book. Suppose we have an $m \times n$ matrix A and a vector $b \in R^m$. If we wish to consider the system $Ax = b$ of m linear equations in n variables and need to emphasize the representation of b as a linear combination of the columns of A, then we let $A_{\bullet j}$ denote the j-th column of A. The equation

$$\sum_{j=1}^{n} A_{\bullet j} x_j = b$$

then gives a *column-oriented representation* of b. We say that column $A_{\bullet j}$ is *used* in the representation of b if the corresponding x_j is nonzero. In the notation $A_{\bullet j}$, the large dot "$\bullet$" represents a "wildcard". This means that all row indices i are included.

In like manner, we denote the i-th row of A by $A_{i\bullet}$. This notation is useful for singling out an individual row from the matrix A which comes up in the row-oriented presentation of the system $Ax = b$. That is, we can use

$$A_{i\bullet} x = b_i \quad \text{as an abbreviation for} \quad \sum_{j=1}^{n} a_{ij} x_j = b_i.$$

By way of illustration, let us consider the matrix

$$A = \begin{bmatrix} 5 & -1 & 0 & 8 \\ 2 & 1 & 2 & 3 \\ 0 & -1 & -1 & 4 \end{bmatrix}.$$

Then we have

$$A_{\bullet 2} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad A_{2\bullet} = [2 \quad 1 \quad 2 \quad 3].$$

Example 2.2: LEAST-NORM SOLUTION OF A LINEAR SYSTEM. Suppose we have a linear system given by the equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, m.$$

Assume the system has a solution. Note that if $n > m$, the linear system could have infinitely many solutions. We seek a solution $x = (x_1, \ldots, x_n)$ having the "least 1-norm," meaning that $|x_1| + \cdots + |x_n|$ is as small as possible among all solutions of the system. So far, the problem is just

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{j=1}^{n} |x_j| \\ \text{subject to} & \displaystyle\sum_{j=1}^{n} A_{\bullet j}x_j = b \end{array} \tag{2.6}$$

where $b = (b_1, \ldots, b_m)$. This is not quite an LP at this point. For one thing, there are no inequalities among the constraints. We can fix that by replacing each variable x_j by the difference of two nonnegative variables: $x_j = u_j - v_j$ for all $j = 1, ..., n$. The constraints then become

$$\begin{gathered} \sum_{j=1}^{n} A_{\bullet j}u_j - \sum_{j=1}^{n} A_{\bullet j}v_j = b \\ u_j \geq 0, \ v_j \geq 0, \quad j = 1, \ldots, n. \end{gathered}$$

How do we—legitimately—get rid of the absolute value signs in (2.6)? It is easy to that if

$$x_j = u_j - v_j, \quad u_j \geq 0, \quad v_j \geq 0, \quad \text{and } u_jv_j = 0, \text{ for all } j = 1, \ldots n,$$

then

$$|x_j| = u_j + v_j, \ j = 1, \ldots, n.$$

The linear program to be solved becomes

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{j=1}^{n} u_j + \sum_{j=1}^{n} v_j \\ \text{subject to} & \displaystyle\sum_{j=1}^{n} A_{\bullet j}u_j - \sum_{j=1}^{n} A_{\bullet j}v_j = b \\ & u_j \geq 0, \ v_j \geq 0, \quad j = 1, \ldots, n. \end{array} \tag{2.7}$$

An appreciation of why the above formulation is valid will be gotten from reading the next section and Exercise 2.8, where you are asked to show that the condition $u_j v_j = 0$ must hold at an optimal extreme point. Based on this assertion and the notation introduced on page 36, we can also express the problem to be solved as the LP

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{j=1}^{n} x_j^+ + \sum_{j=1}^{n} x_j^- \\ \text{subject to} & \displaystyle\sum_{j=1}^{n} A_{\bullet j} x_j^+ - \sum_{j=1}^{n} A_{\bullet j} x_j^- = b \\ & x_j^+ \geq 0,\ x_j^- \geq 0, \quad j = 1, \ldots, n. \end{array} \tag{2.8}$$

Such linear programs come up in signal analysis problems in electrical engineering where estimates of *sparse solutions* of underdetermined linear systems are needed.

Converting equations into inequalities

Sometimes it is necessary to convert an equation (involving two real numbers) into a pair of inequalities. Recall that for real numbers, x and y, the equation $x = y$ is equivalent to the pair of inequalities

$$\begin{aligned} x &\geq y, \\ x &\leq y. \end{aligned}$$

Thus, the equation $\displaystyle\sum_{j=1}^{n} a_{ij} x_j = b_i$ is equivalent to the pair of inequalities

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} x_j &\geq b_i, \\ \sum_{j=1}^{n} a_{ij} x_j &\leq b_i. \end{aligned}$$

Minimization versus maximization

Another general rule is that

$$\max f(x) = -\min\{-f(x)\}. \tag{2.9}$$

Note the effects of the two minus signs in this equation. The function $-f$ is the reflection of f with respect to the horizontal axis. Suppose $\bar{x}$ minimizes $-f(x)$. Then by changing the sign of the function value $-f(\bar{x})$, we find that the maximum value of $f(x)$ is $f(\bar{x})$. The maximum of f and the minimum of $-f$ occur at the same place. The maximum and minimum *values* of these two functions sum to zero. Using this rule, we may convert a maximization problem into a minimization problem, and vice versa.

This is a handy rule when one needs to convert a maximization problem to a minimization problem. For example, if one had an optimization algorithm that solves only minimization problems, it would be possible to use (2.9) to recast any maximization problem to an equivalent minimization problem.

Figure 2.4, given below, depicts the general rule stated in (2.9) between the maxima and minima of a function (solid curve) and the minima and maxima (respectively) of the *negative* of that function (dotted curve).

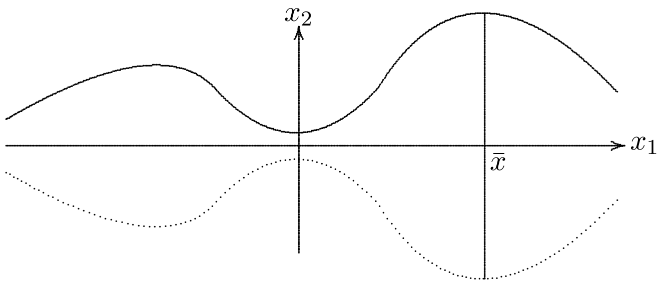

Figure 2.4: Illustration of (2.9).

Some important (LP) language

Consider an LP, say

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}}x \\ \text{subject to} & Ax = b \\ & \quad x \geq 0. \end{array}$$

A vector $\bar{x}$ satisfying the constraints $Ax = b$ and $x \geq 0$ is called a *feasible vector* for this LP. Even if an LP is not in standard form, the analogous definition would apply. A feasible vector for an LP is simply one that satisfies its constraints. In a minimization problem, a feasible vector $\bar{x}$ is *optimal* if $c^{\mathrm{T}}\bar{x} \leq c^{\mathrm{T}}x$ for every feasible vector x.

The feasible region of an LP

The *feasible region* of a linear program is the set of all its feasible vectors. If this set is empty, then the LP is said to be *infeasible*. The feasible region of an LP can be described as the intersection of the solution sets of a finite number of linear inequalities. As a simple illustration of this idea, consider again the linear inequalities

$$\begin{aligned} 3x_1 - 5x_2 &\le 15 \\ 2x_1 + 4x_2 &\le 12 \end{aligned}$$

and Figure 2.2 which depicts their solution set. These linear inequalities could be the constraints of a linear programming problem in which case the doubly shaded region would represent (a portion of) its feasible region.

2.2 Polyhedral sets

Linear equations and linear inequalities can be interpreted geometrically. Let $a \in R^n$ be a nonzero vector and let $b \in R$. Then the solution set of a linear equation, say $a^{\mathrm{T}}x = b$, is called a *hyperplane*. On the other hand, the solutions of the linear inequality $a^{\mathrm{T}}x \le b$ form what is called a *halfspace*, and the same is true for the linear inequality $a^{\mathrm{T}}x \ge b$. For either of these two types of halfspaces, the boundary is a hyperplane, namely the set of x such that $a^{\mathrm{T}}x = b$.

Any subset of R^n that can be represented as the intersection of a finite collection of halfspaces is called a *polyhedron*, or *polyhedral set*. This section covers some properties of polyhedral sets. In Chapter 7 we shall have more to say about polyhedral sets and their structure.

An important consequence of this definition is that the solution set of a linear inequality system is a polyhedron. And since a linear equation is equivalent to a pair of linear inequalities, we can say that the solution set of the following linear system is also polyhedral:

$$\begin{aligned} Fx &\le f \qquad (f \in R^{m_1}) \\ Gx &= g \qquad (g \in R^{m_2}) \\ Hx &\ge h \qquad (h \in R^{m_3}). \end{aligned}$$

Within the class of linear inequality systems, we distinguish those for which the right-hand side vector is zero. A very simple example of such a system is $Ix \geq 0$. A linear inequality system of this form is said to be *homogeneous*.

Let C be the solution set of a homogeneous linear inequality system, say

$$\begin{aligned} Fx &\leq 0 \\ Gx &= 0 \\ Hx &\geq 0. \end{aligned}$$

Then C has the property that

$$x \in C \quad \Longrightarrow \quad \lambda x \in C \quad \text{for all } \lambda \geq 0. \tag{2.10}$$

Any set C for which (2.10) holds is called a *cone*; a polyhedral set for which (2.10) holds is called a *polyhedral cone*.

The feasible region of every linear programming problem is an example of a polyhedron. Indeed, a linear program whose constraints are

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 1 \\ x_j \geq 0, \quad j &= 1, 2, 3 \end{aligned} \tag{2.11}$$

would have a feasible region looking like the shaded tetrahedron in Figure 2.5 below.

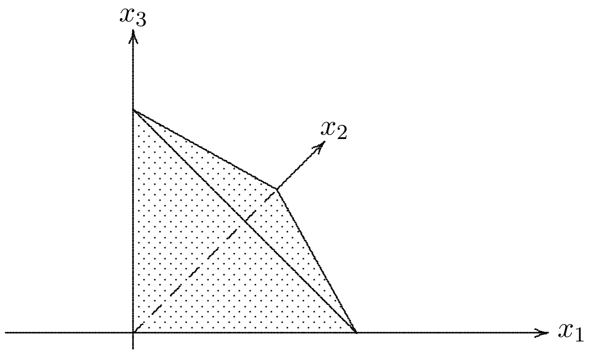

Figure 2.5: Feasible region of (2.11).

The polyhedral region specified by (2.11) is the intersection of four half-spaces. This particular set happens to be *bounded*, that is to say, it is contained within a ball of finite radius. Not every polyhedral set is a bounded,

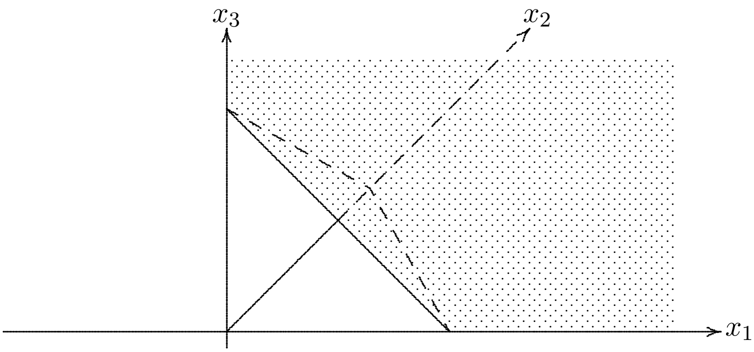

Figure 2.6: Feasible region of (2.12).

however. To illustrate this fact, we revise the example given above by making a small change in the one of the constraints. Consider the polyhedral set given by

$$\begin{aligned} x_1 + x_2 + x_3 &\geq 1 \\ x_j \geq 0, \quad j &= 1, 2, 3. \end{aligned} \tag{2.12}$$

The feasible region corresponding to (2.12) is not bounded as Figure 2.6 is intended to show.

Since a linear equation is equivalent to a pair of linear inequalities, we can regard the feasible region of linear system

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ x_j \geq 0, \quad j &= 1, 2, 3 \end{aligned} \tag{2.13}$$

as a polyhedral set. It is depicted in Figure 2.7.

Convexity of polyhedral sets

A set X (not necessarily polyhedral) is defined to be *convex* if it contains the line segment between any two of its point. The *line segment* between the points x^1 and x^2 in R^n is given by

$$\{x \in R^n : x = \lambda x^1 + (1 - \lambda)x^2,\ 0 \leq \lambda \leq 1\}.$$

Every element of the above set is called a *convex combination* of x^1 and x^2. More generally, if $x^1, \ldots, x^k \in R^n$ and $\lambda_1, \ldots, \lambda_k$ are nonnegative real

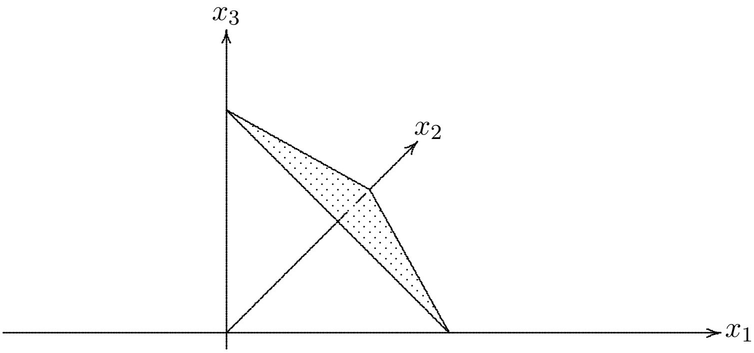

Figure 2.7: Feasible region of (2.13).

numbers satisfying $\lambda_1 + \cdots + \lambda_k = 1$, then

$$\lambda_1 x^1 + \cdots + \lambda_k x^k$$

is called a *convex combination* of $x^1, \ldots, x^k$.

The entire space R^n is clearly a convex set, and so is any halfspace. An elementary property of convex sets is that the intersection of any two of them is again a convex set.[3] It follows, then, that every polyhedron (and hence the feasible region of every linear programming problem) is convex, as is demonstrated in the next paragraph by an elementary argument. Associated with any set S (convex or not), there is another set called its *convex hull* which is defined as the intersection of all convex sets containing S. Accordingly, a set is convex if and only if it equals its convex hull.

There is an alternate definition of a convex hull of the set S. Indeed, the convex hull of S is the set of all points which can be expressed as a convex combination of finitely many points belonging to S. The two definitions of convex hull are equivalent.

The solution sets arising from (2.11), (2.12), and (2.13) are clearly convex. To see that *all* polyhedral sets are convex, consider a linear inequality system given by $Ax \leq b$ where $A \in R^{m \times n}$ and $b \in R^m$. We readily verify that if x^1 and x^2 each satisfy this linear inequality system, then so does $\lambda x^1 + (1-\lambda)x^2$

[3]By convention, the empty set $\emptyset$ is regarded as convex.

for all $\lambda \in [0,1]$[4]. Indeed, for all $\lambda \in [0,1]$ we have

$$A(\lambda x^1 + (1-\lambda)x^2) = \lambda A x^1 + (1-\lambda)Ax^2 \leq \lambda b + (1-\lambda)b = b.$$

The shaded set shown on the left in Figure 2.8 is not convex because it does not contain the line segment between every pair of its elements. (It is a nonconvex polygon.) The set shown on the right in Figure 2.8 is convex, however. This set is polyhedral since it can be viewed as the intersection of a finite collection of halfspaces (in this case five of them).

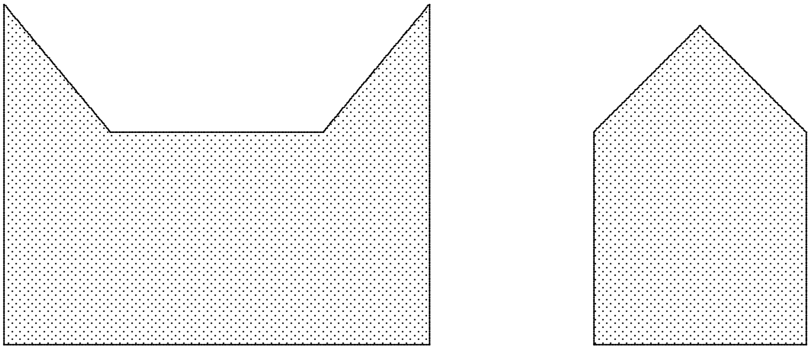

Figure 2.8: Nonconvex polygon (left) and convex polygon (right).

Furthermore, it should be noted that not all convex sets are polyhedral. For instance, disks and ellipsoids are nonpolyhedral convex sets.

Extreme points and their importance in linear programming

Let $\bar{x}$ be an element of a convex set X. We say $\bar{x}$ is an *extreme point* of X if whenever x^1 and x^2 are distinct points in X, then

$$\bar{x} = \lambda x^1 + (1-\lambda)x^2, \quad 0 \leq \lambda \leq 1$$

implies that $\lambda = 0$ or $\lambda = 1$, in which case $\bar{x} = x^1$ or $\bar{x} = x^2$. Thus, $\bar{x}$ does not lie strictly between x^1 and x^2. This is the essence of the extreme point property.

It is a well-established fact (theorem) that

[4] The notation $[0,1]$ means all numbers between 0 and 1 inclusive.

if a linear programming problem in standard form has an optimal solution, then it has an optimal extreme point solution.

This suggests that in trying to find an optimal solution to a linear program in standard form[5], it would be a reasonable idea to confine the search to the (finite) set of extreme points of the feasible region.

Note that this theorem does *not* say that every optimal solution of a linear program is an extreme point. Indeed, there exist linear programs having nonextreme optimal solutions. One such instance is the LP

$$\begin{array}{ll} \text{maximize} & x_2 \\ \text{subject to} & 0 \leq x_1 \leq 1 \\ & 0 \leq x_2 \leq 1 \end{array}$$

whose feasible region is the "unit square." Every point of the form $(x_1, 1)$ with $0 \leq x_1 \leq 1$ is feasible and optimal. Note that any such point lies on an edge between two extreme points.

Linear independence of vectors

Let $v^1, v^2, \ldots, v^k$ be a set of vectors in a finite-dimensional vector space V, such as R^n. The vectors $v^1, v^2, \ldots, v^k$ are *linearly independent* if and only if

$$\alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k = 0$$

implies

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

Another way to put this implication is to say that there is no *nontrivial* linear relationship among the vectors $v^1, v^2, \ldots, v^k$: none of them is a linear combination of the others.

A set of objects possessing some property is *maximal* if the set cannot be enlarged by adjoining another element while at the same time preserving the property. For instance, the set S consisting of the vectors

$$(1, 1, 1), \quad (0, 1, 1), \quad \text{and} \quad (0, 0, 1)$$

[5] The assumption of standard form guarantees that—if it is nonempty—the feasible region of the LP contains an extreme point.

has the property of linear independence. (Taken as an ensemble, the elements of S are linearly independent.) If we form a set of four vectors by adjoining a vector, (b_1, b_2, b_3) to the set S, we no longer have a linearly independent set of vectors. So, with respect to the property of linear independence, the set S is maximal.

A maximal linearly independent set of vectors in a finite-dimensional vector space is called a *basis* for that space. Any set of vectors that properly includes a basis must be *linearly dependent*. For example, when (b_1, b_2, b_3) is adjoined to the set S defined above, we obtain a linearly independent set.

An implication of linear independence

If w is a linear combination of the linearly independent vectors $v^1, v^2, \ldots, v^k$ in the vector space V, that is, if

$$w = \alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k,$$

then the scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$ used in the representation of w are *unique*. To see this, suppose

$$w = \alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k = \beta_1 v^1 + \beta_2 v^2 + \cdots + \beta_k v^k.$$

Then we must have

$$(\alpha_1 - \beta_1)v^1 + (\alpha_2 - \beta_2)v^2 + \cdots + (\alpha_k - \beta_k)v^k = 0.$$

The linear independence of $v^1, v^2, \ldots, v^k$ implies that $\alpha_j - \beta_j = 0$, $j = 1, 2, \ldots, k$. This, in turn, is another way of saying that there is only one way to represent w as a linear combination of $v^1, v^2, \ldots, v^k$.

Matrices and systems of linear equations

When we have a system of m equations in n unknowns (or variables), say

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, \ldots, m, \tag{2.14}$$

the associated matrix $A = [a_{ij}] \in R^{m \times n}$ and vector $b = (b_1, \ldots, b_m) \in R^m$ allow us to express the system as

$$Ax = b. \tag{2.15}$$

The $m \times n$ matrix A gives rise to two special sets of vectors, namely the *rows* $A_{1\bullet}, \dots, A_{m\bullet}$ and the *columns* $A_{\bullet 1}, \dots, A_{\bullet n}$.

In discussing (2.14), we are particularly interested in the columns of A and their relationship to the column vector b. Note that (2.14) has a solution, $\bar{x}$, if and only if the $n+1$ vectors $A_{\bullet 1}, \dots, A_{\bullet n}, b$ are linearly dependent *and*, in the equation expressing linear dependence, there is a nonzero coefficient on b. Indeed, if $\bar{x}$ satisfies (2.14), we have

$$A_{\bullet 1}\bar{x}_1 + \cdots + A_{\bullet n}\bar{x}_n + b(-1) = 0,$$

and since $(\bar{x}_1, \dots, \bar{x}_n, -1)$ is nonzero, the vectors $A_{\bullet 1}, \dots, A_{\bullet n}, b$ are linearly dependent. Conversely, if these vectors are linearly dependent, there exist scalars $\alpha_1, \dots, \alpha_n, \alpha_{n+1}$ not all zero such that

$$A_{\bullet 1}\alpha_1 + \cdots + A_{\bullet n}\alpha_n + b\alpha_{n+1} = 0. \tag{2.16}$$

If $\alpha_{n+1} \neq 0$, then $\bar{x} = -(\alpha_1/\alpha_{n+1}, \dots, \alpha_n/\alpha_{n+1})$ is a solution of (2.15). On the other hand, if $\alpha_{n+1} = 0$, the columns of A are themselves linearly dependent; in such a case, it may happen that there is no solution to (2.15). For instance, if

$$A = \begin{bmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

then $(1, 1, 1, 0)$ would be a nonzero solution of the corresponding equation (2.16), yet clearly (2.15) has no solution since the system

$$\begin{aligned} x_1 + x_2 - 2x_3 &= 1 \\ x_1 + x_2 - 2x_3 &= 0 \end{aligned}$$

is *inconsistent*. However, if the right-hand side vector b had been any multiple of $(1, 1)$, the system would have been *consistent* and hence solvable.

When $b = 0$ the system (2.15) is called *homogeneous*. Such systems always have at least one solution, namely $x = 0$. Solutions to such systems are called *homogeneous solutions*.

The rank of a matrix

The *column rank* of a matrix A is the maximal number of linearly independent columns of A. The *row rank* of a matrix A is the maximal number of linearly independent rows of A. (This equals the column rank of A^{T}, the transpose of A.) From matrix theory we have the following theorem: If A is an $m \times n$ matrix,

$$\text{row rank}(A) = \text{column rank}(A) = \text{rank}(A) \leq \min\{m, n\}.$$

For example, let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \bar{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

The rank of A is 1 whereas the rank of $\bar{A}$ is 2.

There is a procedure called reduction to row-echelon form for computing the rank of a nonzero $m \times n$ matrix A. The idea is to perform elementary row and column operations, including permutations, on A so as to arrive at another matrix $\tilde{A}$ with the property that for some r, $1 \leq r \leq m$,

1. $\tilde{a}_{jj} \neq 0$ for all $j = 1, \ldots, r$
2. $\tilde{a}_{ij} = 0$ if $i > j$ and $j = 1, \ldots, r$
3. $\tilde{a}_{ij} = 0$ if $i > r$ and $j > r$

The integer r turns out to be the rank of the matrix $\tilde{A}$ and the matrix A as well. Notice that the matrix

$$\bar{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

in the example above is already in row-echelon form with $r = 2$. The following matrix is also in row-echelon form:

$$\hat{A} = \begin{bmatrix} 4 & 1 & 0 & -1 \\ 0 & -3 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Notice that in this case the matrix $\hat{A}$ has a row of zeros. Here again $r = 2$.

A theorem of the alternative for systems of linear equations

A system of linear equations, say $Ax = b$, is either consistent or inconsistent. In the latter case, there is a consistent system of linear equations based on the same data. namely

$$\begin{bmatrix} A^{\mathrm{T}} \\ b^{\mathrm{T}} \end{bmatrix} y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.17}$$

The right-hand side of this linear equality system is not uniquely determined. The component 1 can be replaced by any nonzero number, θ, in which case θy would be a solution if y solves (2.17). The italicized assertion below is an example of an "alternative theorem" (also called a "transposition theorem").[6]

Exactly one of the following two systems of equations has a solution.

(i) $\quad Ax = b,$

(ii) $\quad A^{\mathrm{T}}y = 0, \quad b^{\mathrm{T}}y = 1.$

These two systems cannot both have solutions, for otherwise

$$0 = y^{\mathrm{T}}Ax = y^{\mathrm{T}}b = 1$$

which is plainly absurd. That (ii) must have a solution if (i) does not follows easily from the form of system (i) after reduction to row echelon form.

The reduction of the system to row echelon form is based on the reduction of the coefficient matrix A to row echelon form. Indeed, the key idea, as noted by Strang [182, p. 79], is expressed by the theorem that *to any $m \times n$ matrix A, there exists a permutation matrix P, a unit lower triangular L, and an upper trapezoidal matrix U with its nonzero entries in echelon form such that $PA = LU$.* This being so, the system (i) is equivalent to $L^{-1}PAx = Ux = L^{-1}Pb$. The inconsistency of the system (i) will

[6] In the literature, this result has been incorrectly attributed to David Gale in whose 1960 textbook [75, p. 41] it can, indeed, be found. Actually, it goes back at least as far as the doctoral dissertation of T.S. Motzkin published in 1936 where a theorem is proved [139, p. 51] that gives the one stated here as a special case. Versions of the theorem can also be found as exercise 39 in L. Mirsky's 1955 textbook [136, p. 166] and in a 1956 paper [116, p. 222] by H.W. Kuhn discussing solvability and consistency of linear equation systems.

be revealed by the presence of a zero row, say k, in U for which the corresponding entry in $\tilde{b} = L^{-1}Pb$ is nonzero. A solution of system (ii) can then be constructed using this information.

This alternative theorem will be applied in Chapter 11.

Subspaces associated with a matrix

When the system (2.15) has a solution, we say that the columns $A_{\bullet 1}, \ldots, A_{\bullet n}$ *span* the vector b. The set of all vectors $b \in R^m$ spanned by the columns of $A \in R^{m\times n}$ is a vector space (in fact, a *subspace* of R^m). This vector space goes by various names; one of these is *column space* of A. Others are *affine hull* of A or *span* of A. Furthermore, when we think of the linear transformation $x \mapsto Ax$ associated with A, then the column space is just the *range* of this linear transformation, and is often called the *range space.* However it may be called, the object we have in mind is the vector space

$$\mathcal{R}(A) = \{b : b = Ax \text{ for some } x \in R^n\}.$$

As we have already noted, $\mathcal{R}(A)$ is a subspace of R^m. The dimension r of $\mathcal{R}(A)$ is called the *column rank* of A. This is just the cardinality of a maximal linearly independent subset of columns of A. For $A \in R^{m\times n}$, the relationship

$$r = \dim \mathcal{R}(A) \le m$$

always holds.

Another vector space associated with A is the subspace $\mathcal{N}(A)$ of R^n consisting of all vectors that are mapped into the zero vector. Thus

$$\mathcal{N}(A) = \{x : Ax = 0\}.$$

This set is called the *null space* of A. Its elements belong to the *kernel* of the linear transformation $x \mapsto Ax$.

Just as the subspaces $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are defined for the matrix A, so, too, the subspaces $\mathcal{R}(A^{\mathrm{T}})$ and $\mathcal{N}(A^{\mathrm{T}})$ are defined with respect to the matrix A^{T}, the transpose of A.

An important theorem of matrix algebra states that *the column rank of A equals the column rank of A^{T} (or, equivalently, the row rank of A).*

This number, r, is therefore called the *rank* of A, denoted by $\operatorname{rank}(A)$.

Returning to the four subspaces we have mentioned, we can say that

$$\begin{array}{lll} \mathcal{R}(A) & \text{has dimension} & r \\ \mathcal{N}(A) & \text{has dimension} & n-r \\ \mathcal{R}(A^{\mathrm{T}}) & \text{has dimension} & r \\ \mathcal{N}(A^{\mathrm{T}}) & \text{has dimension} & m-r. \end{array}$$

The numbers $n-r$ and $m-r$ are called the *nullity* of A and A^{T}, respectively.

For any subspace V of R^n, there is another subspace called the *orthogonal complement*, $V^{\perp}$, defined as follows:

$$V^{\perp} = \{u : u^{\mathrm{T}}v = 0 \text{ for all } v \in V\}.$$

The subspaces V and $V^{\perp}$ are orthogonal to each other. The dimensions of V and $V^{\perp}$ are always "complementary" in the sense that

$$\dim V + \dim V^{\perp} = n.$$

In the special case of the subspaces associated with the $m \times n$ matrix A, we have

$$\begin{aligned} \mathcal{R}(A) &= (\mathcal{N}(A^{\mathrm{T}}))^{\perp} \\ \mathcal{N}(A) &= (\mathcal{R}(A^{\mathrm{T}}))^{\perp} \\ \mathcal{R}(A^{\mathrm{T}}) &= (\mathcal{N}(A))^{\perp} \\ \mathcal{N}(A^{\mathrm{T}}) &= (\mathcal{R}(A))^{\perp}. \end{aligned}$$

as well as

$$\begin{aligned} R^m &= \mathcal{R}(A) + \mathcal{N}(A^{\mathrm{T}}) \\ R^n &= \mathcal{R}(A^{\mathrm{T}}) + \mathcal{N}(A). \end{aligned}$$

Like any finite-dimensional vector space, the column space of A has a *basis*—a maximal linearly independent subset of vectors that span $\mathcal{R}(A)$. In fact, the basis can be taken as a set of $r = \dim \mathcal{R}(A)$ suitable columns of A itself. For further details see Strang [182, p. 102–114].

Basic solutions of systems of linear equations

In assuming that the matrix A in (2.15) is of order $m \times n$, it is reasonable to assume further that $m \le n$. (Otherwise, the equations in (2.14) are either redundant or inconsistent.) When $r = m = \min\{m, n\}$, we say that A has *full rank*. Note that when A has full rank, it follows that $\mathcal{R}(A) = R^m$ which is a way of saying that the system (2.15) has a solution for every $b \in R^m$. Moreover, A has full rank if and only if there are m columns in A, say $A_{\bullet j_1}, \ldots, A_{\bullet j_m}$, that are linearly independent. Writing

$$B = [A_{\bullet j_1} \ \cdots \ A_{\bullet j_m}],$$

we refer to B as a *basis* in A.

Here is an example to illustrate some of these concepts. Let

$$A = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 1 \\ 0 & -1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{2.18}$$

We leave it as an exercise to show that the first three columns of A are linearly independent, hence

$$B = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} = [A_{\bullet 1} \ A_{\bullet 2} \ A_{\bullet 3}]$$

is a basis in A as are $[A_{\bullet 1} \ A_{\bullet 2} \ A_{\bullet 4}]$ and $[A_{\bullet 1} \ A_{\bullet 3} \ A_{\bullet 4}]$. Observe that $[A_{\bullet 2} \ A_{\bullet 3} \ A_{\bullet 4}]$ is not a basis in A since its columns are linearly dependent as shown by the fact that the equation

$$A_{\bullet 2}x_2 + A_{\bullet 3}x_3 + A_{\bullet 4}x_4 = 0$$

has the nonzero solution $(x_2, x_3, x_4) = (1, 1, -1)$.

Just as it is useful to have a notation $A_{\bullet j}$ for the j-th column of a matrix A, so it is useful to have a notation for a *set* of columns that specify a submatrix of A. A case in point occurs in the designation of a basis. To this end, let $\beta = \{j_1, \ldots, j_m\}$ and write

$$A_{\bullet \beta} = [A_{\bullet j_1} \ \cdots \ A_{\bullet j_m}].$$

Analogously, we could let $\alpha = \{i_1, \dots, i_k\}$ and write

$$A_{\alpha\bullet} = \begin{bmatrix} A_{i_1\bullet} \\ \vdots \\ A_{i_k\bullet} \end{bmatrix}.$$

Putting these two notational devices together, we can define

$$A_{\alpha\beta} = \begin{bmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_m} \\ \vdots & & \vdots \\ a_{i_k j_1} & \cdots & a_{i_k j_m} \end{bmatrix},$$

which is a $k \times m$ *submatrix* of A.

This notational scheme for denoting a submatrix often requires that the index sets α and β be *ordered* in the sense that

$$i_1 < i_2 < \ \cdots \ < i_k \quad \text{and} \quad j_1 < j_2 < \ \cdots \ < j_m.$$

In singling out a particular set of columns of A corresponding to the index set β, we can do the same sort of thing with a vector x writing $x_\beta = (x_{j_1}, \dots, x_{j_m})$ when β is chosen as above. If $x_j = 0$ for all $j \notin \beta$, then (2.15) becomes

$$A_{\bullet\beta} x_\beta = b, \tag{2.19}$$

and when $A_{\bullet\beta}$ is a basis (with inverse $A_{\bullet\beta}^{-1}$), this equation has the unique solution

$$x_\beta = A_{\bullet\beta}^{-1} b.$$

If we let ν denote the set of indices $j \in \{1, \dots, n\}$ such that $j \notin \beta$, we then obtain a submatrix $A_{\bullet\nu}$ and a subvector x_ν such that (2.15) is expressible as

$$A_{\bullet\beta} x_\beta + A_{\bullet\nu} x_\nu = b.$$

When $A_{\bullet\beta}$ is a basis and we set $x_\nu = 0$, we obtain what is called a *basic solution* to (2.15), namely the vector $\bar{x}$ having $\bar{x}_\beta = A_{\bullet\beta}^{-1} b$ and $\bar{x}_\nu = 0$.

With reference to the system of linear equations given by the data in (2.18), we have

Basic index set β	Nonbasic index set ν	Basic solution $\bar{x}$
$\{1,2,3\}$	4	$(3,\ 5,\ 6,\ 0)$
$\{1,2,4\}$	3	$(3, -1,\ 0,\ 6)$
$\{1,3,4\}$	2	$(3,\ 0,\ 1,\ 6)$

(2.20)

When $\bar{x}$ is a solution of a system (2.15), we say that the columns of A *represent* b. The representation of b in terms of A and $\bar{x}$ *uses* column $A_{\bullet j}$ if $\bar{x}_j \neq 0$. Thus, when $\bar{x}$ is a basic solution, it uses only linearly independent columns of A. This can be taken as the definition of a basic solution in cases where the rank of A is less than m (i.e., where the *rows* of A are linearly dependent).

If A is a $m \times n$ matrix of rank m, there are exactly

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!}$$

ways to choose m columns from the n columns of A, or equivalently the index set β from $\{1, 2, \ldots, n\}$. In some cases, the chosen columns maybe linearly dependent, in which case they will not form a basis. For even modest size matrices, this number is extremely large and it is impractical to generate and test every such $m \times m$ submatrix, $A_{\bullet\beta}$.

From what has already been said, it follows that a linear program in standard form can have only finitely many basic feasible solutions. By the same token, there can be only finitely many feasible bases in the matrix A.

Basic feasible solutions of linear programs

Let $\bar{x}$ be a feasible solution of a linear program with constraints $Ax = b$ and $x \geq 0$. The vector $\bar{x}$ is a *basic feasible solution* if the columns $A_{\bullet j}$ associated with all the $\bar{x}_j > 0$ are linearly independent. That is, when $\beta = \{j : \bar{x}_j > 0\}$, then the submatrix $A_{\bullet\beta}$ has linearly independent columns. When A has full rank, m, it will contain at least one $m \times m$ basis $A_{\bullet\beta}$. Relative to the system (2.15), such a matrix is a *feasible basis* if the unique solution of (2.19) is nonnegative. This amounts to saying that $A_{\bullet\beta}$ is a feasible basis if and only if

$$\bar{x}_\beta = A_{\bullet\beta}^{-1} b \geq 0,$$

and as before, the corresponding basic feasible solution (BFS) $\bar{x}$ has

$$\bar{x}_\beta = A_{\bullet\beta}^{-1} b, \qquad \bar{x}_\nu = 0.$$

Note that in some cases, the $\bar{x}$ can be a basic feasible solution without having all the components of $\bar{x}_\beta$ greater than zero; some of the components of $\bar{x}_\beta$ might be zero. A solution of this sort is said to be *degenerate*.

Basic and nonbasic variables

When $A_{\bullet\beta}$ is a basis in A and $\bar{x}$ is a basic solution of (2.15), the x_j with $j \in \beta$ are called *basic variables* (with respect to β), and those with $j \in \nu$ are known as *nonbasic variables.* To illustrate, we recall the data given in (2.18), the basic solutions of which are given in the table (2.20). Notice that only the first and third basic solutions are feasible relative the given system; the second basic solution listed is *not* feasible because it is not a nonnegative vector. It can, however, be said that each of the basic solutions is *nondegenerate* since the components corresponding to the basic columns are nonzero in each case.

Suppose we now change the right-hand side vector b to $\tilde{b} = (1, 1, -1)$. The analog of the table given in (2.20) would then be

Basic index set β	Nonbasic index set ν	Basic solution $\bar{x}$
$\{1, 2, 3\}$	$\{4\}$	$(1, 1, \;\; 0, 0)$
$\{1, 2, 4\}$	$\{3\}$	$(1, 1, \;\; 0, 0)$
$\{1, 3, 4\}$	$\{2\}$	$(1, 0, -1, 1)$

(2.21)

There are a few things to notice about this case. First, because A is the same as before, there is no change in the bases. Second, we see that two different bases give rise to the same degenerate basic feasible solution. Third, we note that one of the basic solutions is infeasible but nondegenerate. With a different right-hand vector, it would be possible to have a basic solution that is infeasible and degenerate at the same time.

Basic feasible solutions and extreme points

Suppose the feasible region is given by the set X:

$$X = \{x : Ax = b, \; x \geq 0\}.$$

The following theorem is of great importance:

A vector $\bar{x}$ is an extreme point of X if and only if it is a basic feasible solution of the constraints $Ax = b, \; x \geq 0$.

This theorem is a useful link between the algebraic notion of basic feasible solution and the intuitively appealing geometric notion of extreme point. Recall the theorem, stated on page 46, that if an LP in standard form has an optimal solution, then it has an optimal extreme point solution. Combining this theorem with the one stated just above, we see that to solve an LP it suffices to consider only basic feasible solutions as candidates for optimal solutions. There are only finitely many bases in A and hence $Ax = b$, $x \geq 0$, can have only finitely many basic feasible solutions. Equivalently, X can have only finitely many extreme points. But, this finite number can be extremely large, even for problems of modest size. For this reason, an algorithm that examines basic feasible solutions must be efficient in selecting them.

2.3 Exercises

2.1 (a) Graph the feasible region of the product-mix problem given in Example 1.3.

(b) On the same graph plot the set of points where the objective function value equals 750. Are any of these points optimal? Why?

(c) Do the same for the objective function value 850. Are any of these points optimal? Why?

2.2 (a) Write the coefficient matrix for the left-hand side of Exercise 1.1.

(b) Are the rows of this matrix linearly independent? Justify your answer.

(c) What is the rank of this matrix?

(d) Write the coefficient matrix for the general problem (with equality constraints, m sources and n destinations).

(e) What is the percentage of nonzero entries to total entries in the matrix given in your answer to (d)? (This is called the *density* of the matrix. When the percentage is low, the matrix is said to be *sparse*. When it is high, the matrix is said to be *dense*.)

2.3 There is a variant of the transportation problem in which the sum of the supplies is allowed to be larger than the sum of the demands, that is,

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j.$$

This is done by adjoining an extra (fictitious) demand

$$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$$

and then allowing each supply location i to "ship" an amount $x_{i,n+1}$ to destination $n+1$ at unit cost $c_{i,n+1} = 0$. Modify Exercise 1.1 by assuming that there is a supply of 600 units at Singapore. Using the technique described above, write the corresponding transportation problem with equality constraints. (Abbreviate the fictitious destination by "FIC.")

2.4 Write down the equations of the (quadrilateral) feasible region of (2.5) depicted in Figure 2.3. Identify all the extreme points of the feasible region by their coordinates and as the intersections of the sides of the feasible region (described, as above, by their equations).

2.5 A nonempty polyhedral convex set has at most a finite number of extreme points. Why is this true?

2.6 The following table gives data for an experiment in which the linear model

$$b = a_1x_1 + a_2x_2 + a_3x_3$$

is postulated.

a_1	a_2	a_3	b
14.92	–19.89	10.66	985.62
17.76	9.16	18.29	846.21
16.45	–23.07	15.77	742.67
15.93	12.04	16.85	780.32
13.99	10.47	19.55	689.62

Write down the linear program for the associated Chebyshev problem. (You need not convert the LP to standard form for this exercise.)

2.7 In general does the linear program for the Chebyshev problem always have a feasible solution? Why?

2.8 Suppose $(u_1, \dots, u_n, v_1, \dots, v)$ is an optimal extreme point solution of the linear program proposed in Example 2.2. Why does $u_jv_j = 0$ hold for all $j = 1, \dots, n$? Explain why solving the LP formulated in Example 2.2 with a method that uses only extreme points of the feasible region will solve the least 1-norm problem stated there.

2.9 In (2.6) the variables x_j are free. Each of these variables can be expressed as the *difference* of two nonnegative variables x_j^+ and x_j^-. The latter variables appear in the objective function of the linear program (2.8) with a plus sign between them. Given that (2.8) will have an optimal basic feasible solution (a fact covered in Chapter 3), explain why this formulation is valid.

2.10 Suppose you are given a system of linear equations given by $Ax = b$ where the matrix A is of order $m \times n$, but you wish to express the solution set of the system as a polyhedron i.e., as the solution set of finitely many linear inequalities. This can be done by replacing each equation $A_{i\bullet}x = b_i$ by two linear inequalities: $A_{i\bullet}x \le b_i$ and $A_{i\bullet}x \ge b_i$, $i = 1, \dots, m$, thereby leading to an equivalent system of $2m$ linear inequalities. Can this be done with fewer than $2m$ linear inequalities? If so, how?

2.11 We have seen that in a linear system with k free variables, each of these free variables can be replaced by the difference of two nonnegative variables. This would add k more variables to the system. Would it be possible to do this with fewer than k more variables? Justify your answer.

2.12 In the text it is asserted that for any (b_1, b_2, b_3) the vectors

$$(1, 1, 1), \quad (0, 1, 1), \quad (0, 0, 1), \quad \text{and} \quad (b_1, b_2, b_3)$$

are linearly dependent. Denote these vectors by v^1, v^2, v^3, v^4, respectively, and find a set of scalar coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ such that

$$\alpha_1 v^1 + \alpha_2 v^2 + \alpha_3 v^3 + \alpha_4 v^4 = 0.$$

How many such sets of coefficients are there in this case?

2.13 Verify the linear independence of columns $A_{\bullet 1}$ $A_{\bullet 2}$, $A_{\bullet 3}$, where A is the matrix given in (2.18).

2.14 A square matrix X of order n whose elements x_{ij} satisfy the linear conditions

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1, \quad i = 1, \dots, n \\ \sum_{i=1}^{n} x_{ij} &= 1, \quad j = 1, \dots, n \\ x_{ij} &\ge 0, \quad i, j = 1, \dots, n \end{aligned}$$

is said to be *doubly stochastic.*

(a) Verify that the set of all doubly stochastic matrices is convex.

(b) What is the order (size) of the matrix of coefficients in the equations through which a doubly stochastic matrix of order n is defined?

(c) Write down a verbose version of the conditions satisfied by a doubly stochastic matrix of order 3.

(d) Explain why the rank of the matrix of coefficients in your answer to (c) is at most 5.

(e) Find a basis in the matrix of coefficients in your answer to (c).

(f) Is the basic feasible solution corresponding to your answer in (e) nondegenerate? Justify your answer.

2.15 Consider the linear program

$$\begin{array}{lrcl} \text{maximize} & 4x_1 + x_2 & & \\ \text{subject to} & x_1 - 3x_2 & \leq & 6 \\ & x_1 + 2x_2 & \leq & 4. \end{array}$$

(a) Plot the feasible region of the above linear program.

(b) List all the extreme points of the feasible region.

(c) Write the equivalent standard form as defined in (1.1).

(d) Show that the extreme points are basic feasible solutions of the LP.

(e) Evaluate the objective function at the extreme points and find the optimal extreme point solution.

(f) Modify the objective function so that the optimal solution is at a different extreme point.

http://www.springer.com/978-1-4939-7053-7

Linear and Nonlinear Optimization
Cottle, R.W.; Thapa, M.N.
2017, XXXI, 614 p. 58 illus., Hardcover
ISBN: 978-1-4939-7053-7