Vectors and Matrices

Appendix A

Vectors and matrices are notational conveniences for dealing with systems of linear equations and inequalities. In particular, they are useful for compactly representing and discussing the linear programming problem:

$$\text{Maximize} \sum_{j=1}^{n} c_j x_j,$$

subject to:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \qquad (i = 1, 2, \ldots, m),$$

$$x_j \geq 0 \qquad (j = 1, 2, \ldots, n).$$

This appendix reviews several properties of vectors and matrices that are especially relevant to this problem. We should note, however, that the material contained here is more technical than is required for understanding the rest of this book. It is included for completeness rather than for background.

A.1 VECTORS

We begin by defining vectors, relations among vectors, and elementary vector operations.

Definition. A *k-dimensional vector* y is an ordered collection of k real numbers $y_1, y_2, \ldots, y_k$, and is written as $y = (y_1, y_2, \ldots, y_k)$. The numbers y_j $(j = 1, 2, \ldots, k)$ are called the *components* of the vector y.

Each of the following are examples of vectors:

i) (1, −3, 0, 5) is a four-dimensional vector. Its first component is 1, its second component is −3, and its third and fourth components are 0 and 5, respectively.
ii) The coefficients $c_1, c_2, \ldots, c_n$ of the linear-programming objective function determine the n-dimensional vector $c = (c_1, c_2, \ldots, c_n)$.
iii) The activity levels $x_1, x_2, \ldots, x_n$ of a linear program define the n-dimensional vector $x = (x_1, x_2, \ldots, x_n)$.
iv) The coefficients $a_{i1}, a_{i2}, \ldots, a_{in}$ of the decision variables in the ith equation of a linear program determine an n-dimensional vector $A^i = (a_{i1}, a_{i2}, \ldots, a_{in})$.
v) The coefficients $a_{1j}, a_{2j}, \ldots, a_{nj}$ of the decision variable x_j in constraints 1 through m of a linear program define an m-dimensional vector which we denote as $A_j = (a_{1j}, a_{2j}, \ldots, a_{mj})$.

Equality and ordering of vectors are defined by comparing the vectors' individual components. Formally, let $y = (y_1, y_2, \ldots, y_k)$ and $z = (z_1, z_2, \ldots, z_k)$ be two k-dimensional vectors. We write:

$$
\begin{array}{llll}
y = z & & \text{when } y_j = z_j & (j = 1, 2, \ldots, k),\\
y \geq z & \text{or} \quad z \leq y & \text{when } y_j \geq z_j & (j = 1, 2, \ldots, k),\\
y > z & \text{or} \quad z < y & \text{when } y_j > z_j & (j = 1, 2, \ldots, k),
\end{array}
$$

and say, respectively, that *y equals z*, y is *greater than or equal to* z and that y is *greater than* z. In the last two cases, we also say that z is *less than or equal to* y and *less than* y. It should be emphasized that *not all vectors* are ordered. For example, if $y = (3, 1, -2)$ and $x = (1, 1, 1)$, then the first two components of y are greater than or equal to the first two components of x but the third component of y is *less* than the corresponding component of x.

A final note: 0 is used to denote the *null vector* (0, 0, …, 0), where the dimension of the vector is understood from context. Thus, if x is a k-dimensional vector, $x \geq 0$ means that each component x_j of the vector x is nonnegative.

We also define scalar multiplication and addition in terms of the components of the vectors.

Definition. *Scalar multiplication* of a vector $y = (y_1, y_2, \ldots, y_k)$ and a scalar α is defined to be a new vector $z = (z_1, z_2, \ldots, z_k)$, written $z = \alpha y$ or $z = y\alpha$, whose components are given by $z_j = \alpha y_j$.

Definition. *Vector addition* of two k-dimensional vectors $x = (x_1, x_2, \ldots, x_k)$ and $y = (y_1, y_2, \ldots, y_k)$ is defined as a new vector $z = (z_1, z_2, \ldots, z_k)$, denoted $z = x+y$, with components given by $z_j = x_j+y_j$.

As an example of scalar multiplication, consider

$$4(3, 0, -1, 8) = (12, 0, -4, 32),$$

and for vector addition,

$$(3, 4, 1, -3) + (1, 3, -2, 5) = (4, 7, -1, 2).$$

Using both operations, we can make the following type of calculation:

$$
\begin{aligned}
(1, 0)x_1 + (0, 1)x_2 + (-3, -8)x_3 &= (x_1, 0) + (0, x_2) + (-3x_3, -8x_3)\\
&= (x_1 - 3x_3, x_2 - 8x_3).
\end{aligned}
$$

It is important to note that y and z must have the same dimensions for vector addition and vector comparisons. Thus $(6, 2, -1) + (4, 0)$ is *not* defined, and $(4, 0, -1) = (4, 0)$ makes *no* sense at all.

A.2 MATRICES

We can now extend these ideas to any rectangular array of numbers, which we call a *matrix*.

Definition. A *matrix* is defined to be a rectangular array of numbers

$$
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n}\\
a_{21} & a_{22} & \cdots & a_{2n}\\
\vdots & & & \vdots\\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix},
$$

whose *dimension* is m by n. A is called *square* if $m = n$. The numbers a_{ij} are referred to as the *elements* of A.

The tableau of a linear programming problem is an example of a matrix.

We define equality of two matrices in terms of their elements just as in the case of vectors.

Definition. Two matrices A and B are said to be *equal*, written $A = B$, if they have the same dimension and their corresponding elements are equal, i.e., $a_{ij} = b_{ij}$ for all i and j.

In some instances it is convenient to think of vectors as merely being *special cases* of matrices. However, we will later prove a number of properties of vectors that do not have straightforward generalizations to matrices.

Definition. A k-by-1 matrix is called a *column vector* and a 1-by-k matrix is called a *row vector*.

The coefficients in row i of the matrix A determine a row vector $A^i = (a_{i1}, a_{i2}, \ldots, a_{in})$, and the coefficients of column j of A determine a column vector $A_j = \langle a_{1j}, a_{2j}, \ldots, a_{mj} \rangle$. For notational convenience, column vectors are frequently written horizontally in angular brackets.

We can define scalar multiplication of a matrix, and addition of two matrices, by the obvious analogs of these definitions for vectors.

Definition. *Scalar multiplication* of a matrix A and a real number α is defined to be a new matrix B, written $B = \alpha A$ or $B = A\alpha$, whose elements b_{ij} are given by $b_{ij} = \alpha a_{ij}$.

For example,

$$3\begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 0 & -9 \end{bmatrix}.$$

Definition. *Addition* of two matrices A and B, both with dimension m by n, is defined as a new matrix C, written $C = A + B$, whose elements c_{ij} are given by $c_{ij} = a_{ij} + b_{ij}$.

For example,

$$\underset{(A)}{\begin{bmatrix} 1 & 2 & 4 \\ 0 & -3 & 1 \end{bmatrix}} + \underset{(B)}{\begin{bmatrix} 2 & 6 & -3 \\ -1 & 4 & 0 \end{bmatrix}} = \underset{(C)}{\begin{bmatrix} 3 & 8 & 1 \\ -1 & 1 & 1 \end{bmatrix}}$$

If two matrices A and B do not have the same dimension, then $A + B$ is undefined.

The product of two matrices can also be defined if the two matrices have appropriate dimensions.

Definition. The *product* of an m-by-p matrix A and a p-by-n matrix B is defined to be a new m-by-n matrix C, written $C = AB$, whose elements c_{ij} are given by:

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

For example,

$$\begin{bmatrix} 1 & 2 \\ 0 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 & -3 \\ 1 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 14 & -3 \\ -3 & -12 & 0 \\ 7 & 22 & -9 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 & 6 & -3 \\ 1 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -7 & -17 \\ 1 & -10 \end{bmatrix}.$$

If the number of columns of A does not equal the number of rows of B, then AB is undefined. Further, from these examples, observe that matrix multiplication is *not commutative*; that is, $AB \neq BA$, in general.

If $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$ is a row vector and $q = \langle q_1, q_2, \ldots, q_m \rangle$ a column vector, then the special case

$$\pi q = \sum_{i=1}^{m} \pi_i q_i$$

of matrix multiplication is sometimes referred to as an *inner product*. It can be visualized by placing the elements of π next to those of q and *adding*, as follows:

$$\begin{aligned}
&\pi_1 \times q_1 = \pi_1 q_1,\\
&\pi_2 \times q_2 = \pi_2 q_2,\\
&\;\vdots \qquad\quad \vdots\\
&\underline{\pi_m \times q_m = \pi_m q_m.}\\
&\pi q = \sum_{i=1}^{m} \pi_i q_i.
\end{aligned}$$

In these terms, the elements c_{ij} of matrix $C = AB$ are found by taking the inner product of A^i (the ith row of A) with B_j (the jth column of B); that is, $c_{ij} = A^i B_j$.

The following properties of matrices can be seen easily by writing out the appropriate expressions in each instance and rearranging the terms:

$$\begin{aligned}
A + B &= B + A && \text{(Commutative law)}\\
A + (B + C) &= (A + B) + C && \text{(Associative law)}\\
A(BC) &= (AB)C && \text{(Associative law)}\\
A(B + C) &= AB + AC && \text{(Distributive law)}
\end{aligned}$$

As a result, $A + B + C$ or ABC is well defined, since the evaluations can be performed in any order.

There are a few special matrices that will be useful in our discussion, so we define them here.

Definition. The *identity matrix* of order m, written I_m (or simply I, when no confusion arises) is a square m-by-m matrix with *ones* along the diagonal and *zeros* elsewhere.

For example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is important to note that for any m-by-m matrix B, $BI_m = I_m B = B$. In particular, $I_m I_m = I_m$ or $II = I$.

Definition. The *transpose* of a matrix A, denoted A^t, is formed by *interchanging* the rows and columns of A; that is, $a^t_{ij} = a_{ji}$.

If

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

then the transpose of A is given by:

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

We can show that $(AB)^t = B^t A^t$ since the ijth element of both sides of the equality is $\sum_k a_{jk} b_{ki}$.

Definition. An *elementary matrix* is a square matrix with one arbitrary column, but otherwise *ones* along the diagonal and *zeros* elsewhere (i.e., an identify matrix with the exception of one column).

For example,

$$E = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 4 & 1 \end{bmatrix}$$

is an elementary matrix.

A.3 LINEAR PROGRAMMING IN MATRIX FORM

The linear-programming problem

Maximize $c_1x_1 + c_2x_2 + \cdots + c_nx_n$,

subject to:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\leq b_1, \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\leq b_2, \\ \vdots \qquad\qquad\qquad \vdots \quad & \quad \vdots \\ a_{1m}x_1 + a_{2m}x_2 + \cdots + a_{mn}x_n &\leq b_m, \\ x_1 \geq 0, \quad x_2 \geq 0, \quad \ldots, \quad x_n &\geq 0, \end{aligned}$$

can now be written in matrix form in a straightforward manner. If we let:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{and} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

be column vectors, the linear system of inequalities is written in matrix form as $Ax \leq b$. Letting $c = (c_1, c_2, \ldots, c_n)$ be a row vector, the objective function is written as cx. Hence,the linear program assumes the following compact form:

Maximize cx,

subject to:

$$Ax \leq b, \qquad x \geq 0.$$

The same problem can also be written in terms of the column vectors A_j of the matrix A as:

Maximize $c_1x_1 + c_2x_2 + \cdots + c_nx_n$,

subject to:

$$\begin{aligned} A_1x_1 + A_2x_2 + \cdots + A_nx_n &\leq b, \\ x_j \geq 0 \qquad (j = 1, 2, \ldots, n)&. \end{aligned}$$

At various times it is convenient to use either of these forms.

The appropriate *dual* linear program is given by:

Minimize $b_1y_1 + b_2y_2 + \cdots + b_my_m$,

subject to:

$$\begin{aligned} a_{11}y_1 + a_{21}y_2 + \cdots + a_{m1}y_m &\geq c_1, \\ a_{12}y_1 + a_{22}y_2 + \cdots + a_{m2}y_m &\geq c_2, \\ \vdots \qquad\qquad\qquad \vdots \quad & \quad \vdots \\ a_{1n}y_1 + a_{2n}y_2 + \cdots + a_{mn}y_m &\geq c_n, \\ y_1 \geq 0, \quad y_2 \geq 0, \quad \ldots, \quad y_m &\geq 0. \end{aligned}$$

Letting

$$y^t = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

be a column vector, since the dual variables are associated with the constraints of the primal problem, we can write the dual linear program in compact form as follows:

Minimize $b^t y^t$,

subject to:

$$A^t y^t \geq c^t, \quad y^t \geq 0.$$

We can also write the dual in terms of the untransposed vectors as follows:

Minimize yb,

subject to:

$$yA \geq c, \quad y \geq 0.$$

In this form it is easy to write the problem in terms of the row vectors A^i of the matrix A, as:

Minimize $y_1 b_1 + y_2 b_2 + \cdots + y_m b_m$,

subject to:

$$y_1 A^1 + y_2 A^2 + \cdots + y_m A^m \geq c,$$
$$y_i \geq 0 \qquad (i = 1, 2, \ldots, m).$$

Finally, we can write the primal and dual problems in equality form. In the primal, we merely define an m-dimensional column vector s measuring the amount of slack in each constraint, and write:

Maximize cx,

subject to:

$$Ax + Is = b,$$
$$x \geq 0, \qquad s \geq 0.$$

In the dual, we define an n-dimensional row vector u measuring the amount of surplus in each dual constraint and write:

Minimize yb,

subject to:

$$yA - uI = c,$$
$$y \geq 0, \qquad u \geq 0.$$

A.4 THE INVERSE OF A MATRIX

Definition. Given a square m-by-m matrix B, if there is an m-by-m matrix D such that

$$DB = BD = I,$$

then D is called the *inverse* of B and is denoted B^{-1}.

Note that B^{-1} does not mean $1/B$ or I/B, since division is *not* defined for matrices. The symbol B^{-1} is just a convenient way to emphasize the relationship between the inverse matrix D and the original matrix B.

There are a number of simple properties of inverses that are sometimes helpful to know.

i) The inverse of a matrix B is unique if it exists.

Proof. Suppose that B^{-1} and A are both inverses of B. Then

$$B^{-1} = IB^{-1} = (AB)B^{-1} = A(BB^{-1}) = A.$$

ii) $I^{-1} = I$ since $II = I$.

iii) If the inverse of A and B exist, then the inverse of AB exists and is given by $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$.

iv) If the inverse of B exists, then the inverse of B^{-1} exists and is given by $(B^{-1})^{-1} = B$.

Proof. $I = I^{-1} = (B^{-1}B)^{-1} = B^{-1}(B^{-1})^{-1}$.

v) If the inverse of B exists, then the inverse of B^t exists and is given by $(B^t)^{-1} = (B^{-1})^t$.

Proof. $I = I^t = (B^{-1}B)^t = B^t(B^{-1})^t$.

The natural question that arises is: Under what circumstances does the inverse of a matrix exist? Consider the square system of equations given by:

$$Bx = Iy = y.$$

If B has an inverse, then multiplying on the left by B^{-1} yields

$$Ix = B^{-1}y,$$

which ''solves'' the original square system of equations for any choice of y. The second system of equations has a unique solution in terms of x for any choice of y, since one variable x_j is isolated in each equation. The first system of equations can be derived from the second by multiplying on the left by B; hence, the two systems are identical in the sense that any $\bar{x}, \bar{y}$ that satisfies one system will also satisfy the other. We can now show that a square matrix B has an inverse if the square system of equations $Bx = y$ has a unique solution x for an arbitrary choice of y.

The solution to this system of equations can be obtained by successively isolating one variable in each equation by a procedure known as *Gauss–Jordan elimination*, which is just the method for solving square systems of equations learned in high-school algebra. Assuming $b_{11} \neq 0$, we can use the first equation to eliminate x_1 from the other equations, giving:

$$\begin{array}{llllll}
x_1 & +\frac{b_{12}}{b_{11}} x_2 + \cdots + & \frac{b_{1m}}{b_{11}} x_m = & \frac{1}{b_{11}} y_1, & & \\
& \left(b_{22} - b_{21}\frac{b_{12}}{b_{11}}\right) x_2 + \cdots + & \left(b_{2m} - b_{21}\frac{b_{1m}}{b_{11}}\right) x_m = & -\frac{b_{21}}{b_{11}} y_1 & + y_2, & \\
& \vdots & & \vdots & \ddots & \\
& \left(b_{m2} - b_{m1}\frac{b_{12}}{b_{11}}\right) x_2 + \cdots + & \left(b_{mm} - b_{m1}\frac{b_{1m}}{b_{11}}\right) x_m = & -\frac{b_{m1}}{b_{11}} y_1 & & + y_m.
\end{array}$$

If $b_{11} = 0$, we merely choose some other variable to isolate in the first equation. In matrix form, the new matrices of the x and y coefficients are given respectively by E_1B and E_1I, where E_1 is an elementary matrix of the form:

$$E_1 = \begin{bmatrix} k_1 & 0 & 0 & \cdots & 0 \\ k_2 & 1 & 0 & \cdots & 0 \\ k_3 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ k_m & 0 & 0 & \cdots & 1 \end{bmatrix}, \qquad \begin{array}{l} k_1 = \frac{1}{b_{11}}, \\ \vdots \\ k_i = -\frac{b_{i1}}{b_{11}} \qquad (i = 2, 3, \ldots, m). \end{array}$$

Further, since b_{11} is chosen to be nonzero, E_1 has an inverse given by:

$$E_1^{-1} = \begin{bmatrix} 1/k_1 & 0 & 0 & \cdots & 0 \\ -k_2 & 1 & 0 & \cdots & 0 \\ -k_3 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ -k_m & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Thus by property (iii) above, if B has an inverse, then E_1B has an inverse and the procedure may be repeated. Some x_j coefficient in the second row of the updated system *must be nonzero*, or no variable can be isolated in the second row, implying that the inverse does not exist. The procedure may be repeated by eliminating this x_j from the other equations. Thus, a new elementary matrix E_2 is defined, and the new system

$$(E_2E_1B)x = (E_2E_1)y$$

has x_1 isolated in equation 1 and x_2 in equation 2.

Repeating the procedure finally gives:

$$(E_mE_{m-1}\cdots E_2E_1B)x = (E_mE_{m-1}\cdots E_2E_1)y$$

with one variable isolated in each equation. If variable x_j is isolated in equation j, the final system reads:

$$\begin{array}{llll} x_1 & & & = \beta_{11}y_1 + \beta_{12}y_2 + \cdots + \beta_{1m}y_m, \\ & x_2 & & = \beta_{21}y_1 + \beta_{22}y_2 + \cdots + \beta_{2m}y_m, \\ & & \ddots & \qquad \vdots \\ & & x_m & = \beta_{m1}y_1 + \beta_{m2}y_2 + \cdots + \beta_{mm}y_m, \end{array}$$

and

$$B^{-1} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & & & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mm} \end{bmatrix}.$$

Equivalently, $B^{-1} = E_mE_{m-1}\cdots E_2E_1$ is expressed in *product form* as the matrix product of elementary matrices. If, at any stage in the procedure, it is not possible to isolate a variable in the row under consideration, then the inverse of the original matrix does not exist.

If x_j has not been isolated in the jth equation, the equations may have to be permuted to determine B^{-1}. This point is illustrated by the following example:

$$B = \begin{bmatrix} \overset{x_1}{0} & \overset{x_2}{2} & \overset{x_3}{4} \\ 2 & 2 & 0 \\ 4 & 6 & 0 \end{bmatrix}. \qquad \text{Isolate } x_2 \text{ in equation 1.}$$

$$E_1 = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -1 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \qquad \overset{E_1B}{\begin{bmatrix} 0 & 1 & 2 \\ 2 & 0 & -4 \\ 4 & 0 & -12 \end{bmatrix}} \;\Bigg|\; \overset{E_1I}{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -1 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}}$$

Isolate x_1 in Eq. 2.

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -2 & 1 \end{bmatrix} \qquad \overset{E_2E_1B}{\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & -2 \\ 0 & 0 & -4 \end{bmatrix}} \;\Bigg|\; \overset{E_2E_1I}{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ -1 & -2 & 1 \end{bmatrix}}$$

Isolate x_3 in Eq. 3.

$$E_3 = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{4} \end{bmatrix} \qquad \overset{E_3E_2E_1B}{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \;\Bigg|\; \overset{E_3E_2E_1I}{\begin{bmatrix} 0 & -1 & \frac{1}{2} \\ 0 & \frac{3}{2} & -\frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix}}.$$

$\uparrow$ x coefficients $\qquad$ $\uparrow$ y coefficients

Rearranging the first and second rows of the last table gives the desired transformation of B into the identity matrix, and shows that:

$$B^{-1} = \begin{bmatrix} 0 & \frac{3}{2} & -\frac{1}{2} \\ 0 & -1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix}.$$

Alternately, if the first and second columns of the last table are interchanged, an identity matrix is produced. Interchanging the first and second columns of B, and performing the same operations as above, has this same effect. Consequently,

$$E_3E_2E_1 = \begin{bmatrix} 0 & -1 & \frac{1}{2} \\ 0 & \frac{3}{2} & -\frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix} \quad \text{is the inverse of} \quad \begin{bmatrix} 2 & 0 & 4 \\ 2 & 2 & 0 \\ 6 & 4 & 0 \end{bmatrix}.$$

In many applications the column order, i.e., the indexing of the variables x_j, is arbitrary, and this last procedure is utilized. That is, one variable is isolated in each row and the variable isolated in row j is considered the jth basic variable (above, the second basic variable would be x_1). Then the product form gives the inverse of the columns to B, reindexed to agree with the ordering of the basic variables.

In computing the inverse of a matrix, it is often helpful to take advantage of any special structure that the matrix may have. To take advantage of this structure, we may *partition* a matrix into a number of smaller matrices, by subdividing its rows and columns.

For example, the matrix A below is partitioned into four submatrices A_{11}, A_{12}, A_{21}, and A_{22}:

$$A = \left[\begin{array}{cc|c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array}\right] = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right].$$

The important point to note is that partitioned matrices obey the usual rules of matrix algebra. For example, multiplication of two partitioned matrices

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}, \qquad \text{and} \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

results in

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \\ A_{31}B_{11} + A_{32}B_{21} & A_{31}B_{12} + A_{32}B_{22} \end{bmatrix},$$

assuming the indicated products are defined; i.e., the matrices A_{ij} and B_{jk} have the appropriate dimensions.

To illustrate that partitioned matrices may be helpful in computing inverses, consider the following example. Let

$$M = \begin{bmatrix} I & Q \\ 0 & R \end{bmatrix},$$

where 0 denotes a matrix with all zero entries. Then

$$M^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

satisfies

$$MM^{-1} = I \quad \text{or} \quad \begin{bmatrix} I & Q \\ 0 & R \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

which implies the following matrix equations:

$$A + QC = I, \qquad B + QD = 0,$$
$$RC = 0, \qquad RD = I.$$

Solving these simultaneous equations gives

$$C = 0, \qquad A = I, \qquad D = R^{-1}, \qquad \text{and} \qquad B = -QR^{-1};$$

or, equivalently,

$$M^{-1} = \begin{bmatrix} I & -QR^{-1} \\ 0 & R^{-1} \end{bmatrix}.$$

Note that we need only compute R^{-1} in order to determine M^{-1} easily. This type of use of partitioned matrices is the essence of many schemes for handling large-scale linear programs with special structures.

A.5 BASES AND REPRESENTATIONS

In Chapters 2, 3, and 4, the concept of a basis plays an important role in developing the computational procedures and fundamental properties of linear programming. In this section, we present the algebraic foundations of this concept.

Definition. *m-dimensional real space* R^m is defined as the collection of all m-dimensional vectors $y = (y_1, y_2, \ldots, y_m)$.

Definition. A set of m-dimensional vectors $A_1, A_2, \ldots, A_k$ is *linearly dependent* if there exist real numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$, *not all zero*, such that

$$\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_k A_k = 0. \tag{1}$$

If the only set of α_j's for which (1) holds is $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, then the m-vectors $A_1, A_2, \ldots, A_k$ are said to be *linearly independent*.

For example, the vectors $(4, 1, 0, -1)$, $(3, 1, 1, -2)$, and $(1, 1, 3, -4)$ are linearly dependent, since

$$2(4, 1, 0, -1) - 3(3, 1, 1, -2) + 1(1, 1, 3, -4) = 0.$$

Further, the *unit* m-dimensional vectors $u_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ for $j = 1, 2, \ldots, m$, with a *plus one* in the jth component and zeros elsewhere, are linearly independent, since

$$\sum_{j=1}^{m} \alpha_j u_j = 0$$

implies that

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

If any of the vectors $A_1, A_2, \ldots, A_k$, say A_r, is the 0 vector (i.e., has all zero components), then, taking $\alpha_r = 1$ and all other $\alpha_j = 0$ shows that the vectors are linearly dependent. Hence, the null vector is linearly dependent on any set of vectors.

Definition. An m-dimensional vector Q is said to be *dependent* on the set of m-dimensional vectors $A_1, A_2, \ldots, A_k$ if Q can be written as a linear combination of these vectors; that is,

$$Q = \lambda_1 A_1 + \lambda_2 A_2 + \cdots + \lambda_k A_k$$

for some real numbers $\lambda_1, \lambda_2, \ldots, \lambda_k$. The k-dimensional vector $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ is said to be the *representation* of Q in terms of $A_1, A_2, \ldots, A_k$.

Note that $(1, 1, 0)$ is not dependent upon $(0, 4, 2)$ and $(0, -1, 3)$, since $\lambda_1(0, 4, 2) + \lambda_2(0, -1, 3) = (0, 4\lambda_1 - \lambda_2, 2\lambda_1 + 3\lambda_2)$ and can never have 1 as its first component.

The m-dimensional vector $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is dependent upon the m-dimensional unit vectors $u_1, u_2, \ldots, u_m$, since

$$(\lambda_1, \lambda_2, \ldots, \lambda_m) = \sum_{j=1}^{m} \lambda_j u_j.$$

Thus, any m-dimensional vector is dependent on the m-dimensional unit vectors. This suggests the following important definition.

Definition. A *basis* of R^m is a set of linearly independent m-dimensional vectors with the property that every vector of R^m is dependent upon these vectors.

Note that the m-dimensional unit vectors $u_1, u_2, \ldots, u_m$ are a basis for R^m, since they are linearly independent and any m-dimensional vector is dependent on them.

We now sketch the proofs of a number of important properties relating bases of real spaces, representations of vectors in terms of bases, changes of bases, and inverses of basis matrices.

Property 1. A set of m-dimensional vectors $A_1, A_2, \ldots, A_r$ is linearly dependent if and only if one of these vectors is dependent upon the others.

Proof. First, suppose that

$$A_r = \sum_{j=1}^{r-1} \lambda_j A_j,$$

so that A_r is dependent upon $A_1, A_2, \ldots, A_{r-1}$. Then, setting $\lambda_r = -1$, we have

$$\sum_{j=1}^{r-1} \lambda_j A_j - \lambda_r A_r = 0,$$

which shows that $A_1, A_2, \ldots, A_r$ are linearly dependent.

Next, if the set of vectors is dependent, then

$$\sum_{j=1}^{r} \alpha_j A_j = 0,$$

with at least one $\alpha_j \neq 0$, say $\alpha_r \neq 0$. Then,

$$A_r = \sum_{j=1}^{r-1} \lambda_j A_j,$$

where

$$\lambda_j = -\left(\frac{\alpha_j}{\alpha_r}\right),$$

and A_r depends upon $A_1, A_2, \ldots, A_{r-1}$. ■

Property 2. The representation of any vector Q in terms of basis vectors $A_1, A_2, \ldots, A_m$ is unique.

Proof. Suppose that Q is represented as both

$$Q = \sum_{j=1}^{m} \lambda_j A_j \quad \text{and} \quad Q = \sum_{j=1}^{m} \lambda'_j A_j.$$

Eliminating Q gives $0 = \sum_{j=1}^{m} (\lambda_j - \lambda'_j) A_j$. Since $A_1, A_2, \ldots, A_m$ constitute a basis, they are linearly independent and each $(\lambda_j - \lambda'_j) = 0$. That is, $\lambda_j = \lambda'_j$, so that the representation must be unique. ■

This proposition actually shows that if Q can be represented in terms of the linearly independent vectors $A_1, A_2, \ldots, A_m$, whether a basis or not, then the representation is unique. If $A_1, A_2, \ldots, A_m$ is a basis, then the representation is always possible because of the definition of a basis.

Several mathematical-programming algorithms, including the simplex method for linear programming, move from one basis to another by introducing a vector into the basis in place of one already there.

Property 3. Let $A_1, A_2, \ldots, A_m$ be a basis for R^m; let $Q \neq 0$ be any m-dimensional vector; and let $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ be the representation of Q in terms of this basis; that is,

$$Q = \sum_{j=1}^{m} \lambda_j A_j. \tag{2}$$

Then, if Q replaces any vector A_r in the basis with $\lambda_r \neq 0$, the new set of vectors is a basis for R^m.

Proof. Suppose that $\lambda_m \neq 0$. First, we show that the vectors $A_1, A_2, \ldots, A_{m-1}, Q$ are linearly independent. Let α_j for $j = 1, 2, \ldots, m$ and α_Q be any real numbers satisfying:

$$\sum_{j=1}^{m-1} \alpha_j A_j + \alpha_Q Q = 0. \tag{3}$$

If $\alpha_Q \neq 0$, then

$$Q = \sum_{j=1}^{m-1} \left(-\frac{\alpha_j}{\alpha_Q}\right) A_j,$$

which with (2) gives two representations of Q in terms of the basis $A_1, A_2, \ldots, A_m$. By Property 2, this is impossible, so $\alpha_Q = 0$. But then, $\alpha_1 = \alpha_2 = \cdots = \alpha_{m-1} = 0$, since $A_1, A_2, \ldots, A_{m-1}$ are linearly independent. Thus, as required, $\alpha_1 = \alpha_2 = \cdots = \alpha_{m-1} = \alpha_Q = 0$ is the only solution to (3).

Second, we show that any m-dimensional vector P can be represented in terms of the vectors $A_1, A_2, \ldots, A_{m-1}, Q$. Since $A_1, A_2, \ldots, A_m$ is a basis, there are constants $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that

$$P = \sum_{j=1}^{m} \alpha_j A_j.$$

Using expression (2) to eliminate A_m, we find that

$$P = \sum_{j=1}^{m-1} \left[\alpha_j - \alpha_m \left(\frac{\lambda_j}{\lambda_m}\right) A_j\right] + \frac{\alpha_m}{\lambda_m} Q,$$

which by definition shows that $A_1, A_2, \ldots, A_{m-1}, Q$ is a basis. ■

Property 4. Let $Q_1, Q_2, \ldots, Q_k$ be a collection of linearly independent m-dimensional vectors, and let $A_1, A_2, \ldots, A_r$ be a basis for R^m. Then $Q_1, Q_2, \ldots, Q_k$ can replace k vectors from $A_1, A_2, \ldots, A_r$ to form a new basis.

Proof. First recall that the 0 vector is not one of the vectors Q_j, since 0 vector is dependent on any set of vectors. For $k = 1$, the result is a consequence of Property 3. The proof is by induction. Suppose, by reindexing if necessary, that $Q_1, Q_2, \ldots, Q_j, A_{j+1}, A_{j+2}, \ldots, A_r$ is a basis. By definition of basis, there are real numbers $\lambda_1, \lambda_2, \ldots, \lambda_r$ such that

$$Q_{j+1} = \lambda_1 Q_1 + \lambda_2 Q_2 + \cdots + \lambda_j Q_j + \lambda_{j+1} A_{j+1} + \lambda_{j+2} A_{j+2} + \cdots + \lambda_r A_r.$$

If $\lambda_i = 0$ for $i = j+1, j+2, \ldots, r$, then Q is represented in terms of $Q_1, Q_2, \ldots, Q_j$, which, by Property 1, contradicts the linear independence of $Q_1, Q_2, \ldots, Q_k$. Thus some, $\lambda_i \neq 0$ for $i = j+1,\ j+2, \ldots, r$, say, $\lambda_{j+1} \neq 0$. By Property 3, then, $Q_1, Q_2, \ldots, Q_{j+1},\ A_{j+1},\ A_{j+2}, \ldots, A_r$ is also a basis. Consequently, whenever $j < k$ of the vectors Q_i can replace j vectors from $A_1, A_2, \ldots, A_r$ to form a basis, $(j+1)$ of them can be used as well, and eventually $Q_1, Q_2, \ldots, Q_k$ can replace k vectors from $A_1, A_2, \ldots, A_r$ to form a basis. ■

Property 5. Every basis for R^m contains m vectors.

Proof. If $Q_1, Q_2, \ldots, Q_k$ and $A_1, A_2, \ldots, A_r$ are two bases, then Property 4 implies that $k \leq r$. By reversing the roles of the Q_j and A_i, we also have $r \leq k$ and thus $k = r$, and every two bases contain the same number of vectors. But the unit m-dimensional vectors $u_1, u_2, \ldots, u_m$ constitute a basis with m-dimensional vectors, and consequently, every basis of R^m must contain m vectors. ■

Property 6. Every collection $Q_1, Q_2, \ldots, Q_k$ of linearly independent m-dimensional vectors is contained in a basis.

Proof. Apply Property 4 with $A_1, A_2, \ldots, A_m$ the unit m-dimensional vectors. ■

Property 7. Every m linearly-independent vectors of R^m form a basis. Every collection of $(m + 1)$ or more vectors in R^m are linearly dependent.

Proof. Immediate, from Properties 5 and 6. ■

If a matrix B is constructed with m linearly-independent column vectors $B_1, B_2, \ldots, B_m$, the properties just developed for vectors are directly related to the concept of a *basis inverse* introduced previously. We will show the relationships by defining the concept of a nonsingular matrix in terms of the independence of its vectors. The usual definition of a nonsingular matrix is that the determinant of the matrix is nonzero. However, this definition stems historically from calculating inverses by the method of cofactors, which is of little computational interest for our purposes and will not be pursued.

Definition. An m-by-m matrix B is said to be *nonsingular* if both its column vectors $B_1, B_2, \ldots, B_m$ and rows vectors $B^1, B^2, \ldots, B^m$ are linearly independent.

Although we will not establish the property here, defining nonsingularity of B merely in terms of linear independence of *either* its column vectors or row vectors is equivalent to this definition. That is, linear independence of either its column or row vectors automatically implies linear independence of the other vectors.

Property 8. An m-by-m matrix B has an inverse if and only if it is nonsingular.

Proof. First, suppose that B has an inverse and that

$$B_1\alpha_1 + B_2\alpha_2 + \cdots + B_m\alpha_m = 0.$$

Letting $\alpha = \langle \alpha_1, \alpha_2, \ldots, \alpha_m \rangle$, in matrix form, this expression says that

$$B\alpha = 0.$$

Thus $(B^{-1})(B\alpha) = B^{-1}(0) = 0$ or $(B^{-1}B)\alpha = I\alpha = \alpha = 0$. That is, $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$, so that vectors $B_1, B_2, \ldots, B_m$ are linearly independent. Similarly, $\alpha B = 0$ implies that

$$\alpha = \alpha(BB^{-1}) = (\alpha B)B^{-1} = 0B^{-1} = 0,$$

so that the rows $B^1, B^2, \ldots, B^m$ are linearly independent.

Next, suppose that $B_1, B_2, \ldots, B_m$ are linearly independent. Then, by Property 7, these vectors are a basis for R^m, so that each unit m-dimensional vector u_j is dependent upon them. That is, for each j,

$$B_1\lambda_1^j + B_2\lambda_2^j + \cdots + B_m\lambda_m^j = u_j \tag{4}$$

for some real numbers $\lambda_1^j, \lambda_2^j, \ldots, \lambda_m^j$. Letting D_j be the column vector $D_j = \langle\lambda_1^j, \lambda_2^j, \ldots, \lambda_m^j\rangle$, Eq. (4) says that

$$BD_j = u_j \quad \text{or} \quad BD = I,$$

where D is a matrix with columns $D_1, D_2, \ldots, D_m$. The same argument applied to the row vectors $B^1, B^2, \ldots, B^m$ shows that there is a matrix D' with $D'B = I$. But $D = ID = (D'B)D = D'(BD) = D'I = D'$, so that $D = D'$ is the inverse of B. ■

Property 8 shows that the rows and columns of a nonsingular matrix inherit properties of bases for R^m and suggests the following definition.

Definition. Let A be an m-by-n matrix and B be any m-by-m submatrix of A. If B is nonsingular, it is called a *basis* for A.

Let B be a basis for A and let A_j be any column of A. Then there is a unique solution $\bar{A}_j = \langle\bar{a}_{1j}, \bar{a}_{2j}, \ldots, \bar{a}_{nj}\rangle$ to the system of equations $B\bar{A}_j = A_j$ given by multiplying both sides of the equality by B^{-1}; that is, $\bar{A} = B^{-1}A_j$. Since

$$B\bar{A}_j = B_1\bar{a}_{1j} + B_2\bar{a}_{2j} + \cdots + B_m\bar{a}_{nj} = A_j,$$

the vector $\bar{A}_j$ is the representation of the column A_j in terms of the basis. Applying Property 3, we see that A_j can replace column B_k to form a new basis if $\bar{a}_{kj} \neq 0$. This result is essential for several mathematical-programming algorithms, including the simplex method for solving linear programs.

A.6 EXTREME POINTS OF LINEAR PROGRAMS

In our discussion of linear programs in the text, we have alluded to the connection between extreme points, or corner points, of feasible regions and basic solutions to linear programs. the material in this section delineates this connection precisely, using concepts of vectors and matrices. In pursuing this objective, this section also indicates why a linear program can always be solved at a basic solution, an insight which adds to our seemingly ad hoc choice of basic feasible solutions in the text as the central focus for the simplex method.

Definition. Let S be a set of points in R^n. A point y in S is called an *extreme point* of S if y cannot be written as $y = \lambda w + (1-\lambda)x$ for two distinct points w and x in S and $0 < \lambda < 1$. That is, y does not lie on the line segment joining any two points of S.

For example, if S is the set of feasible points to the system

$$x_1 + x_2 \leq 6,$$
$$x_2 \leq 3, \quad x_1 \geq 0, \quad x_2 \geq 0,$$

then the extreme points are (0, 0), (0, 3), (3, 3), and (6, 0) (see Fig. (A.1).

The next result interprets the geometric notion of an extreme point for linear programs algebraically in terms of linear independence.

Feasible Extreme Point Theorem. Let S be the set of feasible solutions to the linear program $Ax = b$, $x \geq 0$. Then the feasible point $y = (y_1, y_2, \ldots, y_n)$ is an extreme point of S if and only if the columns of A with $y_i > 0$ are linearly independent.

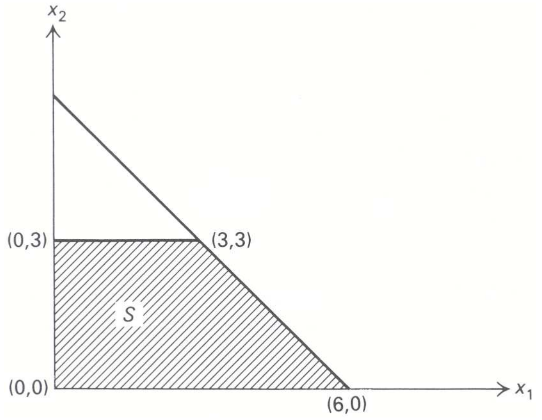

Figure A.1

Proof. By reindexing if necessary, we may assume that only the first r components of y are positive; that is,

$$y_1 > 0, \quad y_2 > 0, \quad \ldots, \quad y_r > 0, \quad y_{r+1} = y_{r+2} = \cdots = y_n = 0.$$

We must show that any vector y solving $Ay = b$, $y \geq 0$, is an extreme point if and only if the first r column $A_1, A_2, \ldots, A_r$ of A are linearly independent. First, suppose that these columns are not linearly independent, so that

$$A_1\alpha_1 + A_2\alpha_2 + \cdots + A_r\alpha_r = 0 \tag{5}$$

for some real numbers $\alpha_1, \alpha_2, \ldots, \alpha_r$ not all zero. If we let x denote the vector $x = (\alpha_1, \alpha_2, \ldots, \alpha_r, 0, \ldots, 0)$, then expression (5) can be written as $Ax = 0$. Now let $w = y + \lambda x$ and $\bar{w} = y - \lambda x$. Then, as long as λ is chosen small enough to satisfy $\lambda|\alpha_j| \leq y_j$ for each component $j = 1, 2, \ldots, r$, both $w \geq 0$ and $\bar{w} \geq 0$. But then, both w and $\bar{w}$ are contained in S, since

$$A(y + \lambda x) = Ay + \lambda Ax = Ay + \lambda(0) = b,$$

and, similarly, $A(y - \lambda x) = b$. However, since $y = \frac{1}{2}(w + \bar{w})$, we see that y is not an extreme point of S in this case. Consequently, every extreme point of S satisfies the linear independence requirement.

Conversely, suppose that $A_1, A_2, \ldots A_r$ are linearly independent. If $y = \lambda w + (1 - \lambda)x$ for some points w and x of S and some $0 < \lambda < 1$, then $y_j = \lambda w_j + (1 - \lambda)x_j$. Since $y_j = 0$ for $j \geq r + 1$ and $w_j \geq 0$, $x_j \geq 0$, then necessarily $w_j = x_j = 0$ for $j \geq r + 1$. Therefore,

$$\begin{aligned} A_1y_1 + A_2y_2 + \cdots + A_ry_r &= A_1w_1 + A_2w_2 + \cdots + A_rw_r \\ &= A_1x_1 + A_2x_2 + \cdots + A_rx_r = b. \end{aligned}$$

Since, by Property 2 in Section A.5, the representation of the vector b in terms of the linearly independent vectors $A_1, A_2, \ldots, A_r$ is unique, then $y_j = z_j = x_j$. Thus the two points w and x cannot be distinct and therefore y is an extreme point of S. ■

If A contains a basis (i.e., the tows of A are linearly independent), then, by Property 6, any collection $A_1, A_2, \ldots, A_r$ of linearly independent vectors can be extended to a basis $A_1, A_2, \ldots, A_m$. The extreme-point theorem shows, in this case, that every extreme point y can be associated with a basic feasible solution, i.e., with a solution satisfying $y_j = 0$ for nonbasic variables y_j, for $j = m + 1, m + 2, \ldots, n$.

Chapter 2 shows that optimal solutions to linear programs can be found at basic feasible solutions or equivalently, now, at extreme points of the feasible region. At this point, let us use the linear-algebra tools of this appendix to drive this result independently. This will motivate the simplex method for solving linear programs algebraically.

Suppose that y is a feasible solution to the linear program

Maximize cx,

subject to:

$$Ax = b, \quad x \geq 0, \tag{6}$$

and, by reindexing variables if necessary, that $y_1 > 0,\ y_2 > 0, \ldots, y_{r+1} > 0$ and $y_{r+2} = y_{r+3} = \cdots = y_n = 0$. If the column A_{r+1} is linearly dependent upon columns $A_1, A_2, \ldots, A_r$, then

$$A_{r+1} = A_1\alpha_1 + A_2\alpha_2 + \cdots + A_r\alpha_r, \tag{7}$$

with at least one of the constants α_j nonzero for $j = 1, 2, \ldots, r$. Multiplying both sides of this expression by θ gives

$$A_{r+1}\theta = A_1(\alpha_1\theta) + A_2(\alpha_2\theta) + \cdots + A_r(\alpha_r\theta), \tag{8}$$

which states that we may simulate the effect of setting $x_{r+1} = \theta$ in (6) by setting $x_1, x_2, \ldots, x_r$, respectively, to $(\alpha_1\theta), (\alpha_2\theta), \ldots, (\alpha_r\theta)$. Taking $\theta = 1$ gives:

$$\tilde{c}_{r+1} = \alpha_1 c_1 + \alpha_2 c_2 + \cdots + \alpha_r c_r$$

as the per-unit profit from the simulated activity of using α_1 units of x_1, α_2 units of x_2, through α_r units of x_r, in place of 1 unit of x_{r+1}.

Letting $\bar{x} = (-\alpha_1, -\alpha_2, \ldots, -\alpha_r, \ +1, 0, \ldots, 0)$, Eq. (8) is rewritten as $A(\theta x) = \theta A\bar{x} = 0$. Here $\bar{x}$ is interpreted as setting x_{r+1} to 1 and decreasing the simulated activity to compensate. Thus,

$$A(y + \theta\bar{x}) = Ay + \theta A\bar{x} = Ay + 0 = b,$$

so that $y + \theta\bar{x}$ is feasible as long as $y + \theta\bar{x} \geq 0$ (this condition is satisfied if θ is chosen so that $|\theta\alpha_j| \leq y_j$ for every component $j = 1, 2, \ldots, r$). The return from $y + \theta\bar{x}$ is given by:

$$c(y + \theta\bar{x}) = cy + \theta c\bar{x} = cy + \theta(c_{r+1} - \tilde{c}_{r+1}).$$

Consequently, if $\tilde{c}_{r+1} < c_{r+1}$, the simulated activity is less profitable than the $(r + 1)$st activity itself, and return improves by increasing θ. If $\tilde{c}_{r+1} > c_{r+1}$, return increases by decreasing θ (i.e., decreasing y_{r+1} and increasing the simulated activity). If $\tilde{c}_{r+1} = c_{r+1}$, return is unaffected by θ.

These observation imply that, if the objective function is bounded from above over the feasible region, then by increasing the simulated activity and decreasing activity y_{r+1}, or vice versa, we can find a new feasible solution whose objective value is at least as large as cy but which contains at least one more zero component than y.

For, suppose that $\tilde{c}_{r+1} \geq c_{r+1}$. Then by decreasing θ from $\theta = 0, c(y + \theta\bar{x}) \geq cy$; eventually $y_j + \theta\bar{x}_j = 0$ for some component $j = 1, 2, \ldots, r + 1$ (possibly $y_{r+1} + \theta\bar{x}_{r+1} = y_{r+1} + \theta = 0$). On the other hand, if $\tilde{c}_{r+1} < c_{r+1}$, then $c(y + \theta\bar{x}) > cy$ as θ increases from $\theta = 0$; if some component of α_j from (7) is positive, then eventually $y_j + \theta\bar{x}_j = y_j - \theta\alpha_j$ reaches 0 as θ increases. (If every $\alpha_j \leq 0$, then we may increase θ indefinitely, $c(y + \theta\bar{x}) \to +\infty$, and the objective value is unbounded over the constraints, contrary to our assumption.) Therefore, if

$$\text{either } \tilde{c}_{r+1} \geq c_{r+1} \quad \text{or} \quad \tilde{c}_{r+1} < c_{r+1},$$

we can find a value for θ such that at least one component of $y_j + \theta\bar{x}_j$ becomes zero for $j = 1, 2, \ldots, r + 1$. Since $y_j = 0$ and $\bar{x}_j = 0$ for $j > r + 1$, $y_j + \theta\bar{x}_j$ remains at 0 for $j > r + 1$. Thus, the entire vector $y + \theta\bar{x}$ contains at least one more positive component than y and $c(y + \theta\bar{x}) \geq cy$.

With a little more argument, we can use this result to show that there must be an optimal extreme-point solution to a linear program.

Optimal Extreme-Point Theorem. If the objective function for a feasible linear program is bounded from above over the feasible region, then there is an optimal solution at an extreme point of the feasible region.

Proof. If y is any feasible solution and the columns A_j of A, with $y_j > 0$, are linearly dependent, then one of these columns depends upon the others (Property 1).

From above, there is a feasible solution x to the linear program with both $cx \geq cy$ and x having one less positive component than y. Either the columns of A with $x_j > 0$ are linearly independent, or the argument may be repeated to find another feasible solution with *one less* positive component. Continuing, we eventually find a feasible solution w with $cw \geq cy$, and the columns of A with $w_j > 0$ are linearly independent. By the *feasible extreme-point theorem*, w is an extreme point of the feasible region.

Consequently, given any feasible point, there is always an extreme point whose objective value is at least as good. Since the number of extreme points is finite (the number of collections of linear independent vectors of A is finite), the extreme point giving the maximum objective value solves the problem. ■