TECHNICAL PAPER
RAC-TP-171
AUGUST 1965

RESEARCH ANALYSIS CORPORATION

Properties of a Generalized Inverse with Applications to Linear Programming Theory

by
Abraham Charnes
Northwestern University

Michael J. L. Kirby
Research Analysis Corporation

RAC

PREPARED FOR THE
DEPARTMENT OF THE ARMY
Contract No. DA 44-188-ARO-1

ADVANCED RESEARCH DIVISION
TECHNICAL PAPER RAC-TP-171
Published August 1965

Properties of a Generalized Inverse with Applications to Linear Programming Theory

by
Abraham Charnes
Northwestern University

Michael J. L. Kirby
Research Analysis Corporation

RESEARCH ANALYSIS CORPORATION
McLEAN, VIRGINIA

FOREWORD

This paper is an expanded version of Office of Naval Research (ONR) Memorandum 91 (Technological Institute, Northwestern University, Evanston, Ill.), which the authors wrote in November 1963. Most of the material in Secs 6 through 8 is also in the original research report. Professor Charnes's work was supported in part by ONR [Research Contract Nonr-1228(10), Project NR 047-021] and the National Institutes of Health (Projects EF 00355-01 and WP 00019-04). Dr. Kirby's work has been conducted in the research program of RAC's Advanced Research Division, under Department of Army sponsorship.

It is hoped that this paper will present an introduction to generalized inverses and some of their applications that will be understandable to any reader with a knowledge of elementary linear algebra. For this reason all theorems in the paper will be proved by using only the basic concepts of linear algebra. In addition, for the sake of completeness, all concepts required in the proofs will be defined and cross-referenced to standard texts.

Nicholas M. Smith
Chief, Advanced Research
Division

CONTENTS

Properties of a Generalized Inverse with Applications to Linear Programming Theory

ABSTRACT

In the first five sections of this paper various properties of a Rao generalized inverse of a matrix are established. A method of computing such an inverse is also given. In order to illustrate the differences between the Rao and other generalized inverses, a survey of results on Penrose-Moore inverses is included.

The last three sections are devoted to showing how a generalized inverse can be used in the theoretical development of the simplex and modified simplex methods of linear programming. In particular, it is shown that the fundamental equations and iteration formulas of these methods can be derived using matrix notation without requiring the assumption that the linear programming problem has no redundant constraints.

1. INTRODUCTION

This paper presents some properties and applications of a Rao generalized inverse $A^{\#}$ of an arbitrary matrix A. After both $A^{\#}$ and the better-known Penrose-Moore inverse A^{+} have been defined and the main results and references on Penrose-Moore inverses have been summarized briefly, A^{+} is shown to be just one of, in general, an infinite number of $A^{\#}$. With several important properties of $A^{\#}$ established, the computational aspect of both the Rao and Penrose-Moore inverses is discussed, and a numerical example shows how much easier it is to compute $A^{\#}$ than A^{+}.

Some uses of $A^{\#}$ in linear programming theory will then be considered. Throughout the literature on linear programming, e.g., in the development of the simplex and modified simplex methods, theorems are proved by using the inverse of a basis matrix for the space spanned by the activity vectors P_j, $j = 1, \ldots, n$ and the stipulations vector P_0. In general, however, such an inverse need not exist. In the distribution and network problems, for example, a basis is not square and hence has no inverse. In fact, for any problem in which redundant constraints exist, a basis matrix will not be square.

Here it is shown that linear programming theorems can be proved just as easily by using a left inverse of a basis matrix as by using the ordinary inverse. It will be shown that such a left inverse always exists and reduces to the regular inverse in the event that the basis matrix is square. It will also be proved that even though the left inverse is not unique it can still be used to give a unique expression for any P_j in terms of the basis. Thus matrix equations of the form $BX_j = P_j$, where B is a basis, can be solved without considering whether B is square.

2. RAO AND PENROSE-MOORE INVERSES

In 1962, C. R. Rao, an Indian statistician, published a paper[1] in which he defined a generalized inverse (g.i.) of a matrix as follows.

Definition 1—A Rao g.i. of an $m \times n$ matrix A is any $n \times m$ matrix $A^{\#}$ such that for any b, for which $AX = b$ is consistent, $X = A^{\#}b$ is a solution.

Rao defined such a g.i. in order to be able to use matrix notation and theory in solving systems of linear equations of the form $AX = b$ in cases where A is rectangular or square and singular.† The question of how to define and use a g.i. for such problems has received attention from time to time in the literature. In particular, attempts have been made to define a g.i. with properties similar

†Rao was particularly interested in using a g.i. for applications to least-squares theory.

to those of an inverse of a nonsingular matrix. The first such attempt was made by Moore.[2,3] The essence of his definition of a g.i. is as follows:

Definition 2—An $n \times m$ matrix A^+ is a g.i. of an $m \times n$ matrix A if AA^+ = projection on the range of A and A^+A = projection on the range of A^+.

Moore established the existence and uniqueness† of A^+ for any A and gave an explicit form for A^+ in terms of the subdeterminants of A and A^T, the transpose of A. Various properties of A^+ and relations between A, A^T, and A^+ were presented by Moore.[3] At the same time a variant of these results was obtained and given an algebraic basis by von Neumann.[4]

Tseng[5,6] extended Moore's results to closed linear operators on a Hilbert space. Further extensions of the work of both von Neumann and Tseng were obtained by Ben-Israel and Charnes.[7,8]

Because the unique notations employed by Moore were not adopted by other mathematicians, his work remained virtually unknown. As a result, other variants of a g.i. were discovered independently by Bjerhammar[9,10] and Penrose.[11,12] The former constructed A^+ by identifying it with a submatrix of the inverse of a particular square nonsingular matrix. He also showed that the general solution of the matrix equation $AX = b$, when consistent, is $X = A^+b + (I - A^+A)Y$, where Y is an arbitrary $n \times 1$ vector. Thus when Y is chosen to be the null vector, it is seen that A^+ also satisfies Definition 1. Hence A^+ is also an $A^\#$.

Meanwhile Penrose defined a g.i. in a manner that can be shown to be equivalent to Definition 2. His definition is stated as follows.

Definition 3—A^+ is the solution of the equations

$$AXA = A$$
$$XAX = X$$
$$(AX)^T = AX$$
$$(XA)^T = XA.$$

Penrose's proof of the existence and uniqueness of A^+ is based on the vanishing of a finite polynomial in A^TA. An extension of his results to the case of infinite matrices was obtained by Ben-Israel and Charnes, as is noted in Ben-Israel and Wersan.[13]

Some other results obtained by Penrose are summarized below.

Theorem 1

(a) $(A^+)^+ = A$.
(b) $(A^T)^+ = (A^+)^T$.
(c) $\det(A) \neq 0$ implies $A^+ = A^{-1}$, the inverse of A.
(d) $(A^TA)^+ = A^+(A^T)^+$ even though in general $(AB)^+ \neq B^+A^+$.
(e) The ranks of A, A^TA, A^+, A^+A are all equal to the trace of A^+A.

Theorem 2

A necessary and sufficient condition for the solvability of $AXB = C$, is that

$$AA^+CB^+B = C,$$

†Uniqueness is the property of A^+ that makes it much more difficult to compute than $A^\#$.

in which case the general solution of $AXB = C$ is

$$X = A^{+}CB^{+} + (Y - A^{+}AYBB^{+})$$

where Y is an arbitrary matrix of the same size as X.

Theorem 2 is valid for any A^{+}, B^{+} that satisfy $AA^{+}A = A$ and $BB^{+}B = B$, respectively. It will be shown subsequently that the condition that A^{+} satisfy $AA^{+}A = A$ is equivalent to Definition 1, so that $AA^{\#}A = A$ for any Rao g.i. $A^{\#}$. Thus Theorem 2 will be seen to hold for any Rao g.i.'s $A^{\#}$, $B^{\#}$.

Moreover it is worth noting that there is a direct relation between the property $AA^{\#}A = A$ and the work of von Neumann[4] on regular rings. In this work, von Neumann defined a ring R as a regular ring if and only if for each element a of R there exists an element x of R such that $axa = a$. An excellent summary of von Neumann's work appears in McCoy (Ref 14, pp 147–49).

Penrose[12] suggested applications of the g.i. in least-squares solutions to inconsistent linear equations. This idea was also used by Tseng,[5,6] whose results suggest the following definition of a g.i.: This definition, also suggested by Penrose (Ref 12, p 18), rests on the least-square character of solutions to linear equations obtained by using A^{+}.

Definition 4—Consider the system of linear equations $AX = b$, where A is $m \times n$. Among all virtual solutions X_b of $AX = b$, defined by

$$||AX_b - b|| = \inf_{X \in E^n} ||AX - b||,$$

where E^n is n-dimensional Euclidean space, there is a unique extremal virtual solution X_b^0, defined by $||X_b^0|| \equiv \inf \{||X_b|| : X_b \text{ satisfies the above relation}\}$. The g.i. A^{+} of A is the matrix corresponding to the linear transformation of b into X_b^0 as b varies in E^m.

This least-square property of $A^{+}b$ was used by Bjerhammar[9,10] in geodetic applications. In particular he used it in adjusting observations that gave rise to singular or ill-conditioned matrices.

Penrose also gave two methods for computing A^{+}, one of which is based on a partition of A that gives an expression for A^{+} in terms of the regular inverses of the partitioned submatrices. The other method is an iterative procedure involving the subdeterminants of $A^{T}A$. Improved methods for computing A^{+} can be found in Ben-Israel and Wersen.[13]

Den Broeder and Charnes[15] gave the following explicit expressions for A^{+} as a limit of a sequence of matrices.

Theorem 3

For any square matrix A, $\lim_{n \to \infty} \sum_{k=1}^{n} A^{T}(I + AA^{T})^{-k}$ exists and

$$A^{+} = \sum_{k=1}^{\infty} A^{T}(I + AA^{T})^{-k},$$

where A^{T} may not be removed from the series as a factor.

Theorem 4

For any square matrix A,

$$A^+ = \lim_{\lambda \to 0} A^T (\lambda \bar{\lambda} I + AA^T)^{-1}.$$

Without loss of generality, only square matrices A need be considered in Theorems 3 and 4. For any $m \times n$ matrix B can be written as a square matrix A by adding the right number of zero rows or columns. Then the $n \times m$ g.i. B^+ is obtained from the corresponding submatrix of A^+.

Results relating A^+ to the principal idempotents of A and the spectral decomposition of A have been obtained by Penrose[11,12] and also by Wedderburn.[16] Hestenes[17,18] also developed a spectral theory for arbitrary $m \times n$ matrices. In doing so, he used A^+ to obtain theorems on structure and some properties of matrices relative to "elementary matrices" and the relations of "*-orthogonality" and "*-commutativity."

Greville[19] gave an iterative procedure for calculating A^+ using successive partitions of A. Moreover (see Ref 20) he gave a very outstanding account of some of Moore's results, following the original Moore approach. Greville also used the least-squares properties of A^+ in regression analysis.

A review of some definitions and applications of A^+ in explicit solutions of systems of linear equations was given by Bjerhammar[21] with numerical examples and statistical applications.

Pyle[22] and Cline[23] considered applications of A^+ to systems of linear equations. Their work in this field originated in Charnes' seminar course at Purdue University in 1955. Pyle used the projections AA^+ and A^+A in a gradient method for solving linear programming problems. These projections were also used by Rosen[24,25] in his conjugate gradient method for solving linear and nonlinear problems.

Bott and Duffin[26] defined an interesting g.i. that is closely related to the von Neumann definition of regularity in a ring.[4] They used their g.i. in the analysis of electrical networks by Maxwell's method.

Ben-Israel and Charnes,[27] following on Bott and Duffin, have used the Penrose-Moore g.i. in the analysis of electrical networks and obtained the explicit solution of a network in terms of its topological and dynamical characteristics. They also pointed out the partially nonoverlapping character of the Bott-Duffin and Penrose-Moore g.i.'s.

Further applications of A^+ have been found by Kalman[28,29] and Florentin[30] in control theory by using its least-squares properties in the mean square error analysis.

More recently A^+ has been used by Charnes, Cooper, and Thompson[31] to resolve questions of the scope and validity of so-called "linear programming under uncertainty" and to characterize optimal stochastic decision rules for n-period problems. Charnes and Kirby[32] use the Rao g.i. in a discussion of the consistency of the constraints of certain chance-constrained programming problems

This completes the brief summary of work done on Penrose-Moore inverses. For a more complete survey of the field the reader is referred to Ben-Israel and Charnes.[8]

3. SOME PROPERTIES OF $A^{\#}$

Now that a brief résumé of the major properties of A^{+} has been presented the derivation of many corresponding properties for the Rao g.i. $A^{\#}$ will be considered. For the sake of completeness the concepts of elementary linear algebra needed in the proofs of the theorems are defined.†

Definition 5—A vector space V is a set of elements called "vectors" satisfying the following axioms.

1. To every pair u and v of vectors of the set there corresponds a vector $v + u$ in such a way that

(a) $v + u = u + v$, i.e., addition is commutative.

(b) $v + (u + w) = (v + u) + w$, i.e., addition is associative.

(c) A unique vector 0 exists such that $v + 0 = v$ for every v in V.

(d) To every vector v a unique vector $-v$ corresponds such that $v + (-v) = 0$.

2. For every scalar α and vector v in the set a vector αv corresponds such that

(a) $\alpha(\beta v) = (\alpha\beta) v$ for any scalar β, i.e., multiplication by scalars is associative.

(b) $1 \times v = v$ for every v.

3. (a) $\alpha(v + u) = \alpha v + \alpha u$, i.e., multiplication by scalars is distributive with respect to addition of vectors.

(b) $(\alpha + \beta)v = \alpha v + \beta v$, i.e., multiplication by vectors is distributive with respect to scalar addition.

Definition 6—A vector subspace L of a vector space V is a collection of vectors in V such that

(a) If $v \in L$ then so is αv for any real scalar α.

(b) If $v, u \in L$, then so is $(u + v) \in L$.

It follows that if $v_1, v_2, \ldots, v_k \in L$ then so is every linear combination of these vectors, i.e., $\Sigma_{j=1}^{k} \alpha_j v_j \in L$ for any set of scalars α_j, $j = 1, \ldots, k$.

Definition 7—A collection of vectors $v_1, \ldots, v_k \in L$ is said to "span" (generate) the subspace L if for every vector $u \in L$ there exists a set of scalars $\alpha_1, \ldots, \alpha_k$ such that $u = \Sigma_{j=1}^{k} \alpha_j v_j$.

Definition 8—A set of vectors $v_1, \ldots, v_k \in L$ is said to be "linearly independent" if $\Sigma_{j=1}^{k} \alpha_j v_j = 0$ implies $\alpha_j = 0$ for all j.

Definition 9—A basis B for a vector subspace L is a set of linearly independent vectors in L that also spans L.

It is clear that the number of vectors in any basis of L is unique. It also follows that if $B \equiv \{v_1, \ldots, v_k\}$ is a basis for L and if u is any vector in L, then u has a unique expression $u = \Sigma_{j=1}^{k} \alpha_j v_j$ as a linear combination of v_j, $j = 1, \ldots, k$, i.e., the values of α_j, $j = 1, \ldots, k$ are unique.

Definition 10—Let A be an $m \times n$ matrix. The "range" of A, $R(A)$, is the set of all $m \times 1$ vectors b such that there exists a solution X to the system of linear equations $AX = b$, i.e., $b \in R(A) \leftrightarrows \exists X \ni AX = b$.

Definition 11—Let A be an $m \times n$ matrix. The null space of A, $N(A)$, is the set of all $n \times 1$ vectors X such that $AX = 0$, i.e., $X \in N(A) \leftrightarrows AX = 0$.

†All theorems will be proved using techniques of elementary linear algebra. Although shorter and more sophisticated proofs are possible, it is more appealing to derive our results from first principles.

It is easy to show that $R(A)$ is a vector subspace of m-dimensional Euclidean space E^m, and $N(A)$ is a subspace of n-dimensional Euclidean space E^n, where E^k is defined, for any integer k, by

$$E^k = \{X : X^T = (x_1, \ldots, x_k),\ x_j \text{ a real number}\}.$$

Definition 12—The rank of a matrix A, written "rank (A)," is the number of vectors in any basis of $R(A)$.

Definition 13—The nullity of a matrix A, written "nullity (A)," is the number of vectors in any basis of $N(A)$.

By Definition 10, $b \in R(A)$ if and only if b can be expressed as a linear combination of the columns of A. Hence it is seen that a definition of rank that is equivalent to Definition 12 is that the rank of A is the maximum number of linearly independent columns of A. This can also be shown to be equal to the maximum number of linearly independent rows of A (Ref 33, p 56).

Theorem 5

Assume that the system of linear equations $AX = b$ is consistent.

Then $A^\#$ is a Rao g.i. of A if and only if $AA^\#A = A$.

Proof: Suppose $AA^\#A = A$.

Then $AX = b$ implies $AA^\#b = AA^\#AX = AX = b$.

Hence $X = A^\#b$ is a solution of $AX = b$, so $A^\#$ is a Rao g.i. by Definition 1.

Conversely suppose $A^\#$ is a Rao g.i. of A.

Let $b = a_j$, the jth column of A.

Then $AX = b = a_j$ is consistent, hence $X = A^\#a_j$ is a solution.

Therefore $a_j = AX = AA^\#a_j$.

Since this must hold for all a_j, $j = 1, \ldots, n$, $A = AA^\#A$.

Using this theorem, it can be seen that the following definition is equivalent to Definition 1.

Definition 14—An $n \times m$ matrix $A^\#$ is a Rao g.i. of an $m \times n$ matrix A if and only if $AA^\#A = A$.

From this definition and Definition 3, it can be seen that A^+ is also a Rao g.i. Thus all the properties of $A^\#$ that will be established for $A^\#$ will also hold for A^+.

The following is required.

Lemma 1.† Let A be an $m \times n$ matrix.

Then n = rank (A) + nullity (A).

If A is $n \times n$ then $R(A)$ and $N(A)$ are both vector subspaces of E^n. Moreover it can be seen from Definitions 10 and 11 that the null vector is the only vector common to $R(A)$ and $N(A)$; yet, every vector $X \in E^n$ must be in one of these two subspaces, as either $AX = b \neq 0$ or $AX = 0$. Hence, except for the 0 vector, $R(A)$ and $N(A)$ partition E^n when A is $n \times n$. This fact will be used below.

In proving Theorem 6 the following will be used.

Definition 15—An $n \times n$ matrix H is said to be "idempotent" if $H^2 = H$.

Lemma 2. Let the $n \times n$ matrix H be idempotent.

Then nullity (H) = rank $(I - H)$, where I is the $n \times n$ identity matrix (i.e., $IX = X$ for every $X \in E^n$).

†A proof of this lemma can be found in Perlis (Ref 33, p 54).

Proof: Let $Y \in R(I - H)$.

Then there exists X such that $(I - H)X = Y$.

Hence $HY = H(I - H)X = (H - H^2)X = 0$, so $Y \in N(H)$.

Conversely, let $Y \in N(H)$ and $Y \neq 0$.

Suppose $Y \in N(I - H)$. Then $(I - H)Y = 0 \Rightarrow Y - HY = Y$ as $Y \in N(H)$. Therefore $Y = 0$ and a contradiction exists.

But, by the remark following Lemma 1, $R(I - H)$ and $N(I - H)$ partition E^n; hence, if $Y \in N(H)$, $Y \neq 0$ then $Y \in R(I - H)$. Since it has been shown that every vector in $N(H)$ is also in $R(I - H)$, and vice versa, the lemma is proved.

Theorem 6

The general solution of the system of homogeneous linear equations $AX = 0$, where A is $m \times n$, is

$$X = (I - A^{\#}A)Y,$$

where Y is an arbitrary $n \times 1$ vector and I is the $n \times n$ identity matrix.

Proof: Let $H = A^{\#}A$.

Since $A = AA^{\#}A = AH$, rank $(A) \leq \min$ [rank (A), rank (H)] (Ref 34, p 75).

Therefore rank $(A) \leq$ rank (H).

Similarly, rank $(H) \leq$ rank (A) as $H = A^{\#}A$.

Therefore rank (A) = rank (H).

Also $H^2 = (A^{\#}A)(A^{\#}A) = A^{\#}(AA^{\#}A) = A^{\#}A = H$; hence H is idempotent and consequently Lemma 2 implies that nullity (H) = rank $(I - H)$. Moreover A is $m \times n$; hence, by Lemma 1, n = rank (A) + nullity (A), so that

$$\begin{aligned} \text{nullity } (A) &= n - \text{rank } (A) \\ &= n - \text{rank } (H) \\ &= \text{nullity } (H). \end{aligned}$$

Therefore

$$\text{nullity } (A) = \text{rank } (I - H). \tag{1}$$

But $A(I - H) = A - AA^{\#}A = 0$. Hence all the columns of $(I - H)$ are in $N(A)$. Moreover Eq 1 implies that $(I - H)$ has the same number of linearly independent columns as any basis for $N(A)$. Hence the columns of $(I - H)$ span $N(A)$. Therefore for any vector $W \in N(A)$ there exists some vector Y such that $W = (I - H)Y$.

Conversely any vector of the form $(I - H)Y$ is in $N(A)$, as $A(I - H)Y = (A - AA^{\#}A)Y = 0$. Thus it has been shown that any vector X is in $N(A)$ (i.e., satisfies $AX = 0$) if and only if there exists some Y such that $X = (I - A^{\#}A)Y$, and so the theorem is proved.

Moreover as $X = A^{\#}b$ is by Definition 1 a particular solution of $AX = b$, whenever $AX = b$ is consistent the following applies:

Theorem 7

The general solution of $AX = b$, when consistent, is

$$X = A^{\#}b + (I - A^{\#}A)Y,$$

where Y is an arbitrary $n \times 1$ matrix.

In most of the preceding results it has been assumed that the system of linear equations $AX = b$ is consistent. The following theorem tells us when this assumption is valid.

Theorem 8

The system of linear equations $AX = b$ is consistent if and only if $AA^{\#}b = b$.

Proof: If $AA^{\#}b = b$, then $X = A^{\#}b$ is a solution of $AX = b$ implying consistency.

Conversely, if $AX = b$ is consistent then $AA^{\#}b = AA^{\#}AX = AX = b$.

Some special cases of $A^{\#}$ are now considered, beginning with the following:

Definition 16—$A^{\#}$ will be called a "right inverse" of A if $AA^{\#} = I$, the $m \times m$ identity matrix.

Definition 17—$A^{\#}$ will be called a "left inverse" of A if $A^{\#}A = I$, the $n \times n$ identity matrix.

It is clear from Definition 14 that any right or left inverse is also a g.i. In Theorem 10 it is proved that the converse is also true; i.e., any g.i. is also a right or left inverse if such an inverse exists.

Theorem 9

Let A be an $m \cdot n$ matrix with $m \leq n$.

Then A has a right inverse if and only if rank $(A) = m$.

In other words, A must have full row rank in order to have a right inverse.

Proof: Suppose there exists an $A^{\#}$ such that $AA^{\#} = I$. Then rank $(AA^{\#})$ = rank $(I) = m$. But rank $(AA^{\#}) = m \leq \min$ [rank (A), rank $(A^{\#})$].

Therefore $m \leq$ rank (A).

But $m \geq$ rank (A) as A is $m \cdot n$; so rank $(A) = m$.

Conversely, if rank $(A) = m$, then $(AA^T)^{-1}$ exists (Ref 35, p 29) and so $A^{\#} = A^T(AA^T)^{-1}$ is a right inverse of A.

Corollary. A necessary and sufficient condition for a matrix to have a left inverse is that it have full column rank.

The remark made following Definition 17 will now be proved.

Theorem 10

Let A be $m \cdot n$ with rank m.

Then $AA^{\#} = I$ for any g.i. $A^{\#}$ of A.

Proof: Let $A^{\#}$ be a g.i. of A.

Then $X = A^{\#}b$ is a solution of $AX = b$ whenever the system is consistent.

But rank $(A) = m$ implies that $AX = b$ is consistent for <u>any</u> $b \in E^m$.†

Hence $AA^{\#}b = b$ for all $b \in E^m$.

Consequently $AA^{\#} = I$, the identity matrix, by definition of I.

Therefore $A^{\#}$ is a right inverse of A.

Corollary. If A has full column rank, then $A^{\#}$ is a g.i. of A if and only if $A^{\#}$ is a left inverse of A.

Theorem 11

Let A be an $m \cdot n$ matrix of rank m.

†Rank $(A) = m$ means that A has m linearly independent columns and so some subset of the columns of A forms a basis for E^m. Hence $R(A)$ is equal to E^m.

Then a necessary and sufficient condition for a g.i. $A^{\#}$ of A to be unique is that $m = n$.

Proof: Let the columns of $A^{\#}$ be a_j, $j = 1, \ldots, n$.

Then $AA^{\#} = I$ implies that $Aa_j = e_j$, where e_j is the $m \times 1$ vector with unity in the jth row and zeros elsewhere.

But the general solution of the system $Aa_j = e_j$ is, by Theorem 7, $a_j = \tilde{A}^{\#} e_j + f$ where f is an arbitrary vector in $N(A)$, and $\tilde{A}^{\#}$ is any g.i. of A.

Hence a_j will be unique only if f is unique.

But if f is unique and $f \in N(A)$, nullity $(A) = 0$.

Since n = rank (A) + nullity (A) by Lemma 1, rank $N(A) = n$.

But, by assumption, rank $(A) = m$.

Therefore a_j, and hence $A^{\#}$, being unique, implies $m = n$.

Conversely, if $m = n$, then rank $(A) = n = m$. By the previous theorem and its corollary, this means that A has both a left and a right inverse; hence $A^{\#} = A^{-1}$ is the unique right inverse of A.

[To see this, let $A_1^{\#}$ and $A_2^{\#}$ be left and right inverses, respectively, of A. Then

$$A_1^{\#} A A_2^{\#} = (A_1^{\#} A) A_2^{\#} = I A_2^{\#} = A_2^{\#}$$

and

$$A_1^{\#} A A_2^{\#} = A_1^{\#} (A A_2^{\#}) = A_1^{\#} I = A_1^{\#}.$$

Therefore

$$A_1^{\#} = A_2^{\#}$$

and the right and left inverses are equal. Since this must hold for all right and left inverses of A, these inverses must be unique. The regular inverse A^{-1} of A is defined as this unique right and left inverse.]

Corollary. If the right inverse of A is unique, then $A^{\#} = A^{-1}$.

Next, the theorem showing how to generate all right inverses of a matrix from any given one is proved.

Theorem 12

Let A be an $m \times n$ matrix with rank m.

Let $AA^{\#} = I$.

Then $A\tilde{A}^{\#} = I$ if and only if there exists an $n \times m$ matrix f such that $\tilde{A}^{\#} = A^{\#} + f$ where each column f_j, $j = 1, \ldots, m$, of f is in $N(A)$.

Proof: Let $f \equiv (f_1, \ldots, f_m)$ be an $n \times m$ matrix such that $f_j \in N(A)$, $j = 1, \ldots, m$.

Let $\tilde{A}^{\#} = A^{\#} + f$.

Then

$$A\tilde{A}^{\#} = A(A^{\#} + f) = I + Af$$
$$= I + (Af_1, \ldots, Af_m)$$
$$= I,$$

as $Af_j = 0$ for all j because $f_j \in N(A)$.

Conversely, if $A\tilde{A}^{\#} = I$, then we can define $f = \tilde{A}^{\#} - A^{\#}$ and $Af = A\tilde{A}^{\#} - AA^{\#} = I - I = 0$. So each column of f is in $N(A)$.

Corollary. Let $A^{\#}A = I$.

Then $\tilde{A}^{\#}A = I$ if and only if $\tilde{A}^{\#} = A^{\#} + f$, where each row of f is a vector in $N(A^T)$.

Theorem 13

Let A be an $m \times n$ matrix of rank n.

Then $AX = b$, $b \neq 0$, has either

(a) no solution if $b \notin R(A)$, or

(b) the unique solution $X = A^{\#}b$ if $b \in R(A)$ and $A^{\#}$ is any left inverse of A.

Proof: (a) This follows directly from the definition of $R(A)$.

(b) Because A has rank n, the system $AX = b$ is a system of m equations exactly n of which are linearly independent. Thus $AX = b$ has a set of n independent equations in n unknowns and so has a unique solution. Since $X = A^{\#}b$ is a solution of $AX = b$, the theorem is proved.

The fact that $A^{\#}b$ is unique regardless of what $A^{\#}$ is chosen can also be seen, as follows: It is known that the general solution of $AX = b$ is $X = A^{\#}b + (I - A^{\#}A)Y$, where Y is arbitrary. But rank $(A) = n$ so that $A^{\#}$ is a left inverse, by the corollary to Theorem 10; hence $I - A^{\#}A = 0$, and the general solution of $AX = b$ is $X = A^{\#}b$.

Let $\tilde{A}^{\#}$ be any other left inverse of A.

Then, by the corollary to Theorem 12, $\tilde{A}^{\#} = A^{\#} + f$, where each row f_i of f is in $N(A^T)$. Therefore

$$\tilde{A}^{\#}b - A^{\#}b = \begin{pmatrix} f_1 b \\ \cdot \\ \cdot \\ f_i b \\ \cdot \\ \cdot \\ f_n b \end{pmatrix}.$$

But $f_i \in N(A^T)$ and $b \in R(A)$; hence $f_i b = 0$, $i = 1, \ldots, n$.†

Therefore $\tilde{A}^{\#}b = A^{\#}b = X$ is unique as $\tilde{A}^{\#}$ was any other left inverse of A.

Theorem 14

Let A be an $m \times n$ matrix.

Let rank $A = n$. Then (a) the equation $\omega^T A = c^T$ has the solution $\omega^T = c^T A^{\#}$, where $A^{\#}$ is any left inverse of A; and (b) ω^T is unique if and only if $n = m$.

Proof: (a) If $\omega^T = c^T A^{\#}$, then $\omega^T A = c^T A^{\#} A = c^T$. Hence $\omega^T = c^T A^{\#}$ is a solution of $\omega^T A = c^T$.

(b) The general solution of $\omega^T A = c^T$ is $\omega^T = c^T A^{\#} + f$, where $f \in N(A^T)$. Therefore ω^T is unique if and only if f is unique.

But f is unique if and only if nullity $(A^T) = 0$, and nullity $(A^T) = 0$ if and only if rank $(A) = m$.

Therefore ω^T is unique if and only if $n = m$.

†This follows from the fact that $b \in R(A) \rightarrow \exists X$, $AX = b$ and $f_i \in N(A) \rightarrow f_i A = 0$; hence $f_i b = f_i AX = 0X = 0$.

4. COMPUTATION OF $A^{\#}$

In order to avoid unnecessary complications in the algebraic treatment in this section, A will be made a square matrix by adding rows or columns of zeros, whichever is necessary. The necessary change in the system of linear equation $AX = b$ is to extend b by a number of zeros if rows of zeros are added to A, and to extend X by a number of variables if columns of zeros are added to A. These changes do not in any way alter the given system of equations since adding rows results in adding the identity $0 = 0$ to the given system, whereas adding columns always makes the new variables in X appear in the system with the coefficient 0.

Moreover the g.i. for the original matrix can always be obtained from the g.i. of the extended square matrix by dropping some rows or columns. For if A, the original matrix, has $m \leq n$, and

$$\hat{A} \begin{pmatrix} A \\ 0_1 \end{pmatrix}$$

is the extended matrix, where 0_1 is the $(n - m) \times n$ matrix of zeros, then $\hat{A}\hat{A}^{\#}\hat{A} = \hat{A}$ means that

$$\begin{pmatrix} A \\ 0_1 \end{pmatrix} = \begin{pmatrix} A \\ 0_1 \end{pmatrix} \begin{pmatrix} Q_1 Q_2 \end{pmatrix} \begin{pmatrix} A \\ 0_1 \end{pmatrix} \quad \begin{pmatrix} AQ_1 AQ_2 \\ 0_1 \quad 0_2 \end{pmatrix} \begin{pmatrix} A \\ 0_1 \end{pmatrix} = \begin{pmatrix} AQ_1 A \\ 0_1 \end{pmatrix},$$

where $\hat{A}^{\#} \equiv (Q_1, Q_2)$, Q_1 is $n \times m$, Q_2 is $n \times (n - m)$, and 0_1 and 0_2 are matrices all the elements of which are 0.

But

$$\begin{pmatrix} A \\ 0_1 \end{pmatrix} = \begin{pmatrix} AQ_1 A \\ 0_1 \end{pmatrix}.$$

means that $A = AQ_1A$. Hence Q_1 is a g.i. of A and is obtained from $\hat{A}^{\#}$ by dropping Q_2, the last $n - m$ columns of $\hat{A}^{\#}$.

Similarly, if A has $m > n$, $m - n$ columns of zeros are added to A to get the extended matrix $\hat{A}$ and $A^{\#}$ is obtained by dropping the last $m - n$ rows from $\hat{A}^{\#}$.

Thus the system $AX = b$ where A is square will be considered without loss of generality.

Definition 18—The following matrix operations are called "elementary row operations":

(a) interchange of two rows;

(b) multiplication of a row by a non-zero scalar; and

(c) replacement of the i th row by the sum of the i th row and k times the j th row, where $i \neq j$ and k is any scalar.

Definition 19—A matrix H is said to be "row-equivalent" to a matrix A if H can be obtained by performing a succession of elementary row operations on A.

From Perlis (Ref 33, p 42) we have the result that for any square matrix A there exists a matrix H that is row-equivalent to A and such that:

(a) all diagonal elements of H are zeros or ones;

(b) when the i th diagonal element is 0, all elements in the i th row are 0 and all elements in the ith column below the diagonal element are 0; and

(c) when the ith diagonal element is 1, the ith column is a unit vector and all elements preceding the ith in the ith row are 0.

Moreover, since each elementary row operation is the result of premultiplying A by a nonsingular matrix, it follows from Perlis that there exists an $n \times n$ nonsingular matrix Q such that $H = QA$ (see also Ref 36, p 24). This leads to the following theorem:

Theorem 15

Let A, H, Q be as defined above.

Then (i) H is idempotent.

(ii) $AH = A$.

(iii) $AQA = A$.

Hence Q is a Rao g.i. of A.

Proof: (i) This follows directly from properties a, b, and c following Definition 19. For if h_i, the ith row of H, has 0 as its diagonal element, then the matrix $D = HH \equiv (d_{ij})$ has $h_i h^j \equiv d_{ij} = 0$, $j = 1, \ldots, n$, since h_i is a zero vector where h^j is the jth column of H. If h^j has 1 as its diagonal element it is a unit vector; hence $h_i h^j \equiv d_{ij} = h_{ij}$, $i = 1, \ldots, n$, the ijth element of H. Since, by a, every diagonal element of H is 0 or 1, $D = H$ and so i is proved.

(ii) Let $H = QA$ where Q is nonsingular.

Hence

$$A = Q^{-1} H$$

so that

$$AH = Q^{-1} H^2 = Q^{-1} H \text{ by i}$$
$$= A.$$

(iii) Since $QA = H$, $AQA = AH = A$ by ii.

Thus by Definition 14, Q is a Rao g.i. of A.

Since Q is the product of all the matrices that give the elementary row operations involved in the reduction of A to H, Q can be found by performing, in turn, exactly the same elementary row operations on the $n \times n$ identity matrix I that we perform on A in reducing it to H. In other words, beginning with A and I, and applying the method of sweep-out and interchange of rows, if necessary, to bring unities to the diagonal, to reduce A to H, (i.e., using the Gaussian elimination technique) Q will be the matrix obtained by performing in order exactly the same operations on I that are performed on A. Thus a g.i. considered above is computed in the same way as a regular inverse when it exists.

5. EXAMPLE OF $A^{\#}$ AND A'

The technique for computing $A^{\#}$, discussed in the preceding section, is illustrated here. The computation may be abridged to a large extent by omitting some intermediate steps, but it is presented here in full to illustrate the method. Let

$$A = \begin{pmatrix} 2 & 1 & -1 & 1 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

Then

Extended matrix				Identity matrix			
(2)	1	−1	1	1	0	0	0
1	0	−1	1	0	1	0	0
1	1	0	0	0	0	1	0
0	0	0	0	0	0	0	1
1	½	−½	½	½	0	0	0
0	−½	−½	½	−½	1	0	0
0	(½)	½	−½	−½	0	1	0
0	0	0	0	0	0	0	1
1	0	−1	1	1	0	−1	0
0	0	0	0	−1	1	1	0
0	1	1	−1	−1	0	2	0
0	0	0	0	0	0	0	1
1	0	−1	1	1	0	−1	0
0	1	1	−1	−1	0	2	0
0	0	0	0	−1	1	1	0
0	0	0	0	0	0	0	1
Matrix H				Matrix Q			

So that

$$A^{\#} \begin{pmatrix} 1 & 0 & -1 \\ -1 & 0 & 2 \\ -1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

In these computations the second block was obtained by taking the 2 (circled in the first block) as a pivot and then, by successive subtraction, sweeping out the first column. The third block was obtained by pivoting on the ½ (circled in the second block) and sweeping out the second column. The last block is obtained simply by interchanging third-block rows 2 and 3.

It is easily verified that $A^{\#}$ given above is a g.i. inverse of A. $A^{\#}$ is not, however, unique. For example, another g.i. of A is

$$\hat{A}^{\#} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ -1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The computation of A^{+} is presented below in order to show the increased computation required in such an operation. This method of finding A^{+} is contained in Ref 13.

The formula for A^{+} is

$$A^{+} = \begin{pmatrix} I_2 \\ A^{\prime} \end{pmatrix} \left(\left(I_2 + AA^{\prime} \right)^{-1} 0_{2,2} \right) EA^{\prime}.$$

$0_{2 \times 2}$ is the 2×2 null matrix. Δ and E are defined by Ben-Israel and Wersan.[13]

$$A^{\prime}A = \begin{pmatrix} 6 & 3 & -3 & 3 \\ 3 & 2 & -1 & 1 \\ -3 & -1 & 2 & -2 \\ 3 & 1 & -2 & 2 \end{pmatrix}$$

Sweep-out techniques are then used on A^TA and A^T to get

A^TA				A^T		
(6)	3	−3	3	2	1	1
3	2	−1	1	1	0	1
−3	−1	2	−2	−1	−1	0
3	1	−2	2	1	1	0
1	½	−½	½	⅓	⅙	⅙
0	(½)	½	−½	0	−½	½
0	½	½	−½	0	−½	½
0	−½	−½	½	0	½	−½
1	0	−1	1	⅓	⅔	−⅓
0	1	1	−1	0	−1	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0

The second block was obtained by pivoting on the 6 (circled) and then sweeping out the first column. The last block was obtained by pivoting on the ½ (circled) and sweeping out the second column.

Then, by the definitions of Δ and EA^T in the Ben-Israel and Wersan paper,[13]

$$\Lambda \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \text{ and } EA^T = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence

$$\Lambda\Lambda^T = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

and

$$I_2 + \Lambda\Lambda^T = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}.$$

Therefore

$$(I_2 + \Lambda\Lambda^T)^{-1} = \frac{1}{5}\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

and

$$A^+ = \frac{1}{5}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 3 & 2 & 0 & 0 \\ 2 & 3 & 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \frac{1}{15}\begin{pmatrix} 3 & 0 & 3 \\ 2 & -5 & 7 \\ -1 & -5 & 4 \\ 1 & 5 & -4 \end{pmatrix}.$$

It can be verified that A^+ does indeed satisfy Definition 3. Thus the great increase in computation that results when A^+ is computed instead of $A^\#$ can be

seen. Since the linear-programming results developed below hold for $A^{\#}$ as well as A^{+}, it will be more efficient to use $A^{\#}$ in solving linear programming problems.

6. THE SIMPLEX ALGORITHM

Consider the following linear programming problem:
maximize

$$c^T \lambda$$

subject to

$$P\lambda = P_0,$$
$$\lambda \geq 0$$

where P is an $m \times n$ matrix and the columns of P are P_j, $j = 1, \ldots, n$.

Let L be the subspace of E^m that is spanned by the vectors $(P_1, \ldots, P_n)$. Thus L is the set of all linear combinations of P_j, $j = 1, \ldots, n$.

Let B be a basis for L. Then each vector in B is in L; and B has, at most, m elements, as L is subspace of E^m. It will be assumed that B has s elements. Thus B is an $m \times s$ matrix with $s \leq m$, and because of the linear independence of the vectors in a basis, rank $(B) = s$.

Let $B = \{P_i : i \in I\}$ and renumber P_j, $j = 1, \ldots, n$ so that $B = (P_1, \ldots, P_s)$.

Then, by definition of a basis, any P_j, $j = 1, \ldots, n$ can be expressed uniquely as a linear combination of the basis vectors as follows:

$$P_j = \sum_{i \in I} P_i x_{ij} = B X_j, \qquad j = 1, \ldots, n, \tag{2}$$

where

$$X_j^T = (x_{1j}, \ldots, x_{sj}).$$

Suppose the basis is changed by removing P_r from B and inserting P_k in its place. Then, provided $x_{rk} \neq 0$, the set of vectors $\{P_i : i \in I, i \neq r\}$ and P_k will form another basis for L.

This follows from the easily proved Lemma 3.

Lemma 3. Let L be the vector space spanned by $\{u_1, u_2, \ldots, u_n\}$. Let $W_1 \in L$ be given by

$$W_1 = \lambda_1 u_1 + \sum_{i=2}^{n} \lambda_i u_i \text{ with } \lambda_1 \neq 0.$$

Then (a) the space spanned by $\{W_1, u_2, \ldots, u_n\}$ is L and

(b) if $\{u_1, u_2, \ldots, u_n\}$ are linearly independent (i.e., form a basis for L) then so are $\{W_1, u_2, \ldots, u_n\}$ (i.e., the latter set is also a basis for L).

This new basis will be denoted by $\hat{B}$. That is, $\hat{B} = (P_1, \ldots, P_{r-1}, P_k, P_{r+1}, \ldots,)$.

Define

$$\eta^T = [-(x_{1k}/x_{rk}), \ldots, -(x_{r-1,k}/x_{rk}), 1/x_{rk}, -(x_{r+1,k}/x_{rk}), \ldots, -(x_{sk}/x_{rk})].$$

From Eq 2 we have

$$P_k = \sum_{\substack{i=1 \\ i \neq r}} P_i(x_{ik}) + P_r(x_{rk}).$$

Hence

$$P_r = \sum_{\substack{i=1 \\ i \neq r}} P_i\left(-\frac{x_{ik}}{x_{rk}}\right) + \frac{P_k}{x_{rk}}$$

or, in matrix notation

$$P_r = \hat{B}\eta \tag{3}$$

and

$$P_i = \hat{B}e_i, \; i=1, \; i \neq r, \tag{4}$$

where e_i is the $s \times 1$ vector with $+1$ in the ith row and zeros elsewhere.

Let E be the $s \times s$ matrix defined by $E \equiv (e_1, \ldots, e_{r-1}, \eta, e_{r+1}, \ldots, e_s)$. Then E is nonsingular, from the definitions of η and e_i.†
Also from Eqs 3 and 4

$$B = (P_1, \ldots, P_r, \ldots, P_s) = (\hat{B}e_1, \ldots, \hat{B}\eta, \ldots, \hat{B}e_s) = \hat{B}E.$$

But B has full column rank and hence, by the corollary to Theorem 10, $B^{\#}$ is a left inverse.
Therefore

$$B^{\#}B = I = B^{\#}\hat{B}E$$

or

$$IE^{-1} = E^{-1} = B^{\#}\hat{B}EE^{-1} = B^{\#}\hat{B}.$$

Hence

$$EE^{-1} = I = EB^{\#}\hat{B}$$

so that a left inverse of the new basis $\hat{B}$ is given by $EB^{\#}$, i.e.,

$$\hat{B}^{\#} = EB^{\#}. \tag{5}$$

Definition 20—Given the system $AX = b$, where A is $m \times n$ and rank $(A) = s$, and any $m \times s$ basis B whose columns are columns of A, then the solution X_B of the system $BX_B = b$ is called the "basic feasible solution" for the system $AX = b$ with basis B.

Now suppose that the basic feasible solution corresponding to the basis B is λ_B. Then $B\lambda_B = P_0$; hence

$$\lambda_B = B^{\#}P_0 \tag{6}$$

and, by Theorem 13, λ_B is unique.

†This is true because the rth diagonal element of E is $1/x_{rk} \neq 0$, and all columns of E other than the rth are unit vectors.

Also because B is a basis, it is known from Eq 2 that $BX_j = P_j$, $j = 1, \ldots, n$, and hence

$$X_j = B^{\#}P_j, \qquad j = 1, \ldots, n. \tag{7}$$

Furthermore the X_j are unique.

Then for the new basis $\hat{B}$ it follows that $\hat{\lambda}_B = \hat{B}^{\#}P_0$ and $\hat{X}_j = \hat{B}^{\#}P_j$, $j = 1, \ldots, n$ are the unique values of $\hat{\lambda}_B$ and $\hat{X}_j$. Hence, using Eq 5.

$$\hat{\lambda}_B = EB^{\#}P_0 = E\lambda_B \tag{8}$$

and

$$\hat{X}_j = EB^{\#}P_j = EX_j, \qquad j = 1, \ldots, n. \tag{9}$$

Equations 8 and 9 represent the simplex iteration formula of linear programming theory because, if Eq 9 is put in component form,

$$\hat{x}_{ij} = \sum_{\substack{r=1 \\ r \neq k}} e_{ir}x_{rj} + \eta x_{kj}$$

or

$$\hat{x}_{ij} = x_{ij} + \left(-\frac{x_{ik}}{x_{rk}}\right)(x_{rj}).$$

Therefore

$$\hat{x}_{ij} = x_{ij} - x_{ik}\left(\frac{x_{rj}}{x_{rk}}\right), \qquad \begin{matrix} j = 1, \ldots, n \\ i = 1, \ldots, s, \end{matrix}$$

which is the ordinary equation used in the simplex algorithm (Ref 37, Vol I, p 142).

Similarly

$$\hat{\lambda}_j = \lambda_j - x_{ik}\left(\frac{\lambda_r}{x_{rk}}\right), \qquad j = 1, \ldots, s.$$

Thus it has been shown that given any basis B and the corresponding simplex tableau, whose columns are given by Eqs 6 and 7, all succeeding tableaus computed using the simplex algorithm will be the same as those obtained by computing $B^{\#}$ at each iteration. Hence the simplex algorithm works perfectly well when the basis is not square.

From simplex theory it is also known that for an optimal basis B the $m \times 1$ vector of dual variables ω is the solution of the system of equations

$$\omega^T B = c_B^T, \text{ where } c_B^T = [c_1, \ldots, c_m]. \tag{10}$$

This is true because, from Eq 10, $\omega^T B\lambda_B = c_B^T\lambda_B$ = optimal value of objective function = z, where λ_B is the optimal basic feasible solution. From the fact that $B\lambda_B = P_0$, $\omega^T P_0 = z$ which, by the dual theorem (Ref 37, Vol I, Chap.VI), means that ω is optimal for the dual of our given problem.

But Eq 10 says that $\omega^T = c_B^T B^{\#}$, which, by Theorem 14, is unique if and only if B is square. Thus it has been shown that in a distribution model the

optimal solution of the dual is not unique, since the model has one redundant constraint; hence B is not square.

7. THE MODIFIED SIMPLEX METHOD

The linear programming problem given at the start of Section 6 will be reconsidered here. The fundamental equation of the modified simplex method (Ref 37, Vol II, pp 471–74) is

$$\begin{bmatrix} P_j \\ m\times 1 \\ -c_j \\ 1\times 1 \end{bmatrix} \quad \begin{bmatrix} B & 0 \\ m\times s & m\times 1 \\ -c_B^T & 1 \\ 1\times s & 1\times 1 \end{bmatrix} \begin{bmatrix} X_j \\ s\times 1 \\ z_j - c_j \\ 1\times 1 \end{bmatrix}, \tag{11}$$

where P_j is any column of P, B is a basis for the space spanned by the columns of P, X_j is the expression of P_j in terms of the basis, c_B is the vector of costs of the $P_i \in B$, and z_j is defined by $z_j \equiv c_B^T X_j$.

Let $B^\#$ be any left inverse of B. Then if the $m\times 1$ vector ω is defined by $\omega^T B \equiv c_B^T$ (i.e., $\omega^T = c_B^T B^\#$)

$$\begin{bmatrix} B^\# & 0 \\ \omega^T & 1 \end{bmatrix} \begin{bmatrix} B & 0 \\ -c_B^T & 1 \end{bmatrix} \quad \begin{bmatrix} I & 0 \\ \omega^T B - c_B^T & 1 \end{bmatrix} = \begin{bmatrix} I & 0 \\ s\times s & s\times 1 \\ 0 & 1 \\ 1\times s & 1\times 1 \end{bmatrix}.$$

Using Eq 11 the following is obtained:

$$\begin{bmatrix} X_j \\ z_j - c_j \end{bmatrix} \quad \begin{bmatrix} B^\# & 0 \\ \omega^T & 1 \end{bmatrix} \begin{bmatrix} P_j \\ -c_j \end{bmatrix}, \tag{12}$$

where $\omega^T = c_B^T B^\#$.

The fundamental Eq 11 can be generalized to the whole tableau by defining $X_0 \equiv \lambda$ and $z_0 - c_0 \equiv z \equiv$ value of objective function and $c_0 \equiv 0$.

Thus the entire $(s + 1) \times (n + 1)$ simplex tableau, where the $z_j - c_j$ appear in the $s + 1$st row, can be generated by Eq 13:

$$\begin{bmatrix} X_0, X_1, \ldots, X_n \\ z, z_1 - c_1, \ldots, z_n - c_n \end{bmatrix} \quad \begin{bmatrix} B^\# & 0 \\ s\times m & s\times 1 \\ c_B^T B^\# & 1 \\ 1\times 1 & 1\times 1 \end{bmatrix} \begin{bmatrix} P_0, P_1, \ldots, P_n \\ 0, -c_1, \ldots, -c_n \end{bmatrix}. \tag{13}$$

Consequently given any $m\times s$ matrix B, which is a basis for the space spanned by the columns of P, $B^\#$ can be computed and hence the entire simplex tableau, including the $z_j - c_j$ row, by using Eq 13. Since it has already been

shown that the calculation of $\hat{B}''$ from B'' can be done by using the usual simplex formulas, it can be concluded that the modified simplex method will work for any basis, regardless of whether the basis is square.

8. EXAMPLE

To illustrate the theory the following network problem will be solved. An arbitrary basis B will be selected and computed and Eq 12 will be used to generate the initial simplex tableau. The simplex algorithm is then used until an optimal solution is reached. Finally a left inverse of the optimal basis is computed and used to find the optimal vector of dual variables and hence the optimal value of the dual objective function. Diagrammatically the network is as shown in Fig. 1.

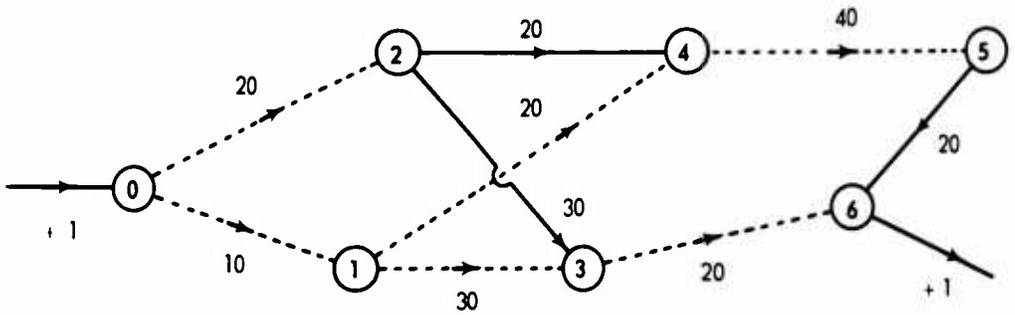


Fig. 1—Diagram of the Network Problem

Arrow shows direction of link; number beside arrow is the c_{ij} of the link. Dotted arrow indicates link used to form initial basis.

The problem to be solved is:
maximize

$$10x_{01} + 20x_{02} + 30x_{13} + 20x_{14} + 30x_{23} + 20x_{24} + 20x_{36} + 40x_{45} + 20x_{56}$$

subject to

$$\begin{array}{rrrrrrrrrrr}
-x_{01} & -x_{02} & & & & & & & & = & -1 \\
x_{01} & & - x_{13} & - x_{14} & & & & & & = & 0 \\
& x_{02} & & & -x_{23} & - x_{24} & & & & = & 0 \\
& & x_{13} & & + x_{23} & & - x_{36} & & & = & 0 \\
& & & x_{14} & & + x_{24} & & - x_{45} & & = & 0 \\
& & & & & & & x_{45} & - x_{56} & = & 0 \\
& & & & & & x_{36} & & + x_{56} & = & 1
\end{array} \qquad (14)$$

$$x_{ij} \geq 0, \text{ all } i, j.$$

where x_{ij} is the amount programmed to flow from node i to node j.

In the incidence matrix for the problem, "+" and "−" are used to denote "+1" and "−1" and all blanks represent 0.

Node	c_{ij}									
		10	20	30	30	20	20	20	40	20
	Links									
		0–1	0–2	1–3	2–3	1–4	2–4	3–6	4–5	5–6
0	−	−	−							
1		+		−		−				
2			+		−		−			
3				+	+			−		
4						+	+		−	
5									+	−
6	+							+		+
P_j	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9

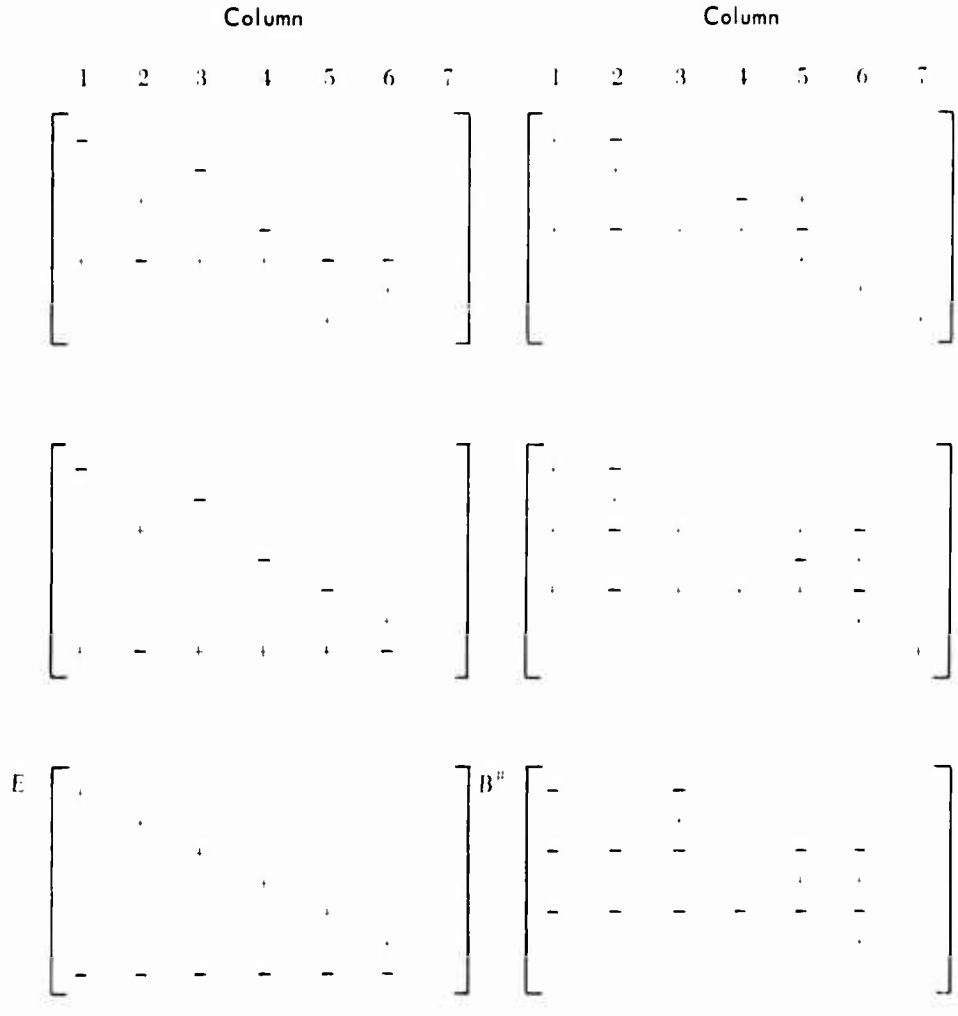
The initial basis matrix is $B = (P_1, P_2, P_3, P_5, P_7, P_8)$. $B^{\#}$ is computed as follows:

Column 1 2 3 4 5 6 7

$$\begin{bmatrix} - & - & & & & & \\ + & & - & - & & & \\ & + & & & & & \\ & & + & & - & & \\ & & & + & & - & \\ & & & & & + & \\ & & & & + & & \end{bmatrix} \quad \begin{bmatrix} + & & & & & & \\ & + & & & & & \\ & & + & & & & \\ & & & + & & & \\ & & & & + & & \\ & & & & & + & \\ & & & & & & + \end{bmatrix}$$

$$\begin{bmatrix} - & & & & & & \\ + & - & - & - & & & \\ & + & & & & & \\ & & + & & - & & \\ & & & + & & - & \\ & & & & & + & \\ & & & & + & & \end{bmatrix} \quad \begin{bmatrix} + & - & & & & & \\ & + & & & & & \\ & & + & & & & \\ & & & + & & & \\ & & & & + & & \\ & & & & & + & \\ & & & & & & + \end{bmatrix}$$

$$\begin{bmatrix} - & & & & & & \\ & & - & & & & \\ + & & & & & & \\ & + & & - & - & & \\ & & + & & & & \\ & & & + & & - & \\ & & & & + & & \end{bmatrix} \quad \begin{bmatrix} + & - & & & & & \\ & + & & & & & \\ + & & - & + & - & & \\ & & & + & & & \\ & & & & + & & \\ & & & & & + & \\ & & & & & & + \end{bmatrix}$$



From this choice of B,

$$c_B^I = (c_{01}, c_{02}, c_{13}, c_{14}, c_{36}, c_{45}) = (10, 20, 30, 20, 20, 40).$$

Thus

$$\omega^I = c_B^I B^\# = (-60, -50, -40, -20, -30, 10, 0).$$

Equation 13 is then used to generate the initial simplex tableau as follows:

$$\begin{bmatrix} B^\# & 0 \\ \omega^I & 1 \end{bmatrix} \begin{bmatrix} P_0 & P_1 \cdots P_9 \\ 0 & -c_1 \cdots -c_9 \end{bmatrix} =$$

c_B	Basis vector	c_j									
		0	10	20	30	30	20	20	20	40	20
		x_j									
		X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
10	P_1	+	+			+		+			
20	P_2			+		−		−			
30	P_3	+			+	+					+
20	P_5						+	+			−
20	P_7	+							+		+
40	P_8									+	−
	$z_j - c_j$	60				−10		−10			−30

These simplex iterations then lead to the following optimal tableau:
First iteration: Insert P_9 into the basis and remove P_7.
Second iteration: Insert P_4 into the basis and remove P_3.
Third iteration: Insert P_6 into the basis and remove P_5.
Optimal tableau:

c_B	Basis vector	x_j									
		X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
10	P_1		+		−		−				
20	P_2	+		+	+		+				
30	P_4				+	+			−		
20	P_6	+					+	+	+		
20	P_9	+							+		+
40	P_8	+							+	+	
	$z_j - c_j$	100			10		10		30		

Thus the optimal solution is: $x_{02} = +1$, $x_{24} = +1$, $x_{45} = +1$, $x_{56} = +1$; and $z = 100$ is the optimal value of the objective function. Since the optimal basis $B^* = (P_1, P_2, P_4, P_6, P_9, P_8)$ we find that

$$B^{*\#} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{bmatrix} & + & & & & & \\ - & - & & & & & \\ & & & + & & & \\ - & - & - & - & & & \\ - & - & - & - & - & - & \\ - & - & - & - & - & & \end{bmatrix} \end{array}$$

so that ω^{*T}, the optimal dual solution, is given by

$$\omega^{*T} = c_B^{T*} B^{*\#} = (10, 20, 30, 20, 20, 40)B^{*\#}$$
$$(-100, -90, -80, -50, -60, -20, 0).$$

Hence $\omega^{*T} P_0 = 100$ = optimal value of dual objective function. Thus the values of the objective functions of the primal and dual problems are equal at the optimum.

REFERENCES

CITED REFERENCES

1. C. R. Rao, "A Note on a Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics," J. Roy. Stat. Soc.,24 Series B: 152–58 (1962).
2. E. H. Moore, Bull. Amer. Math. Soc.,26: 394–95 (1920).
3. ——, "General Analysis, Part 1," Memoirs of the American Philosophical Society.1, (1935).
4. J. von Neumann, "On Regular Rings," Proc. Natl. Acad. Sci.,22: 707–13 (1936).
5. Yu. Ya. Tseng, "Virtual Solutions and General Inversions," Uspehi Mat. Nauk (N.S.), 11: 213–15 (1956).
6. ——, "Sur les Solutions des Equations Operatrices Fonctionnelles entre les Espaces Unitaires. Solutions Extremales. Solutions Virtuelles." C. R. Acad. Sci. Paris, 228: 640–41 (1949).
7. A. Ben-Israel and A. Charnes, "Projection Properties and Neumann-Euler Expansions for the Moore-Penrose Inverse of an Arbitrary Matrix," ONR Research Memo. 40; Northwestern University, the Technological Institute, Evanston, Ill. (1961).
8. ——, ——, "Contributions to the Theory of Generalized Inverses," J. Soc. Indus. and Appl. Math., 11 (3): 667–99 (Sep 63).
9. A. Bjerhammar, "Application of Calculus of Matrices to Method of Least Squares; with Special Reference to Geodetic Calculations," Trans. Roy. Inst. Tech. Stockholm, 49: 1–86 (1951).
10. ——, "Rectangular Reciprocal Matrices with Special Reference to Geodetic Calculations," Bull. Geodesique, pp 188–220 (1951).
11. R. Penrose, "A Generalized Inverse for Matrices," Proc. Cambridge Philos. Soc., 51: 406–13 (1955).
12. ——, "On Best Approximate Solution of Linear Matrix Equations," Proc. Cambridge Philos. Soc., 52: 17–19 (1956).
13. A. Ben-Israel and S. J. Wersan, "A Least Squares Method for Computing the Generalized Inverse of an Arbitrary Matrix," ONR Research Memo. 61,Northwestern University, the Technological Institute, Evanston, Ill. (1962).
14. N. H. McCoy, Rings and Ideals, Carus Mathematical Monographs, No. 8, Amer. Math. Assoc., Buffalo, N. Y., pp 147–49 (1948).
15. G. G. Den Broeder, Jr. and A. Charnes, "Contributions to the Theory of a Generalized Inverse for Matrices," ONR Research Memo. 39.Northwestern University, the Technological Institute, Evanston, Ill., Jan 62.
16. J. H. M. Wedderburn, Lectures on Matrices, Amer. Math. Soc. Colloquium Publications,New York, 1934, Vol XVIII.
17. M. R. Hestenes, "Inversion of Matrices by Biorthogonalization and Related Results," J. Soc. Indus. Appl. Math.,6 (1): 54–90 (1958).
18. ——, "Relative Hermitian Matrices." Pacific J. Math.,1: 225–45 (1961).
19. T. N. E. Greville, "Some Applications of the Pseudo Inverse of a Matrix," J. Soc. Indus. Appl. Math., Rev. 2, (1): 15–22 (Jan 60).
20. ——, "The Pseudo Inverse of a Rectangular or Singular Matrix and its Applications to the Solution of Systems of Linear Equations," J. Soc. Indus. Appl. Math., Rev. 1: 38–43 (1959).

21. A. Bjerhammar, A Generalized Matrix Algebra, Kungl. Tekn. Hogsk. Handl., Stockholm, 1958.
22. L. D. Pyle, "The Generalized Inverse in Linear Programming," Ph.D. thesis, Purdue University, Lafayette, Ind., Jun 60.
23. R. E. Cline, "Representations for the Generalized Inverse of Matrices with Applications in Linear Programming," Ph.D. thesis, Purdue University, Lafayette, Ind., Jun 63.
24. J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part I: Linear Constraints," J. Soc. Indus. Appl. Math., 8 (1): 181–217 (Mar 60).
25. ——, "The Gradient Projection Method for Nonlinear Programming, Part II: Nonlinear Constraints," J. Soc. Indus. Appl. Math., 9 (4): 514–32 (Dec 61).
26. R. Bott and R. J. Duffin, "On the Algebra of Networks," Trans. Amer. Math. Soc., 74 (1): 99–109 (1953).
27. A. Ben-Israel and A. Charnes, "Generalized Inverses and the Bott-Duffin Network Analysis," J. Math. Anal. Appl., 7 (3): 428–35 (Dec 63).
28. R. E. Kalman, "Contributions to the Theory of Optimal Control," Bol. Soc. Mat. Mexicana, pp 102–19 (1960).
29. ——, "New Methods and Results in Linear Prediction and Filtering Theory," RIAS, Tech. Rept. 61-1, Baltimore, 1961.
30. J. J. Florentin, "Optimal Control of Continuous Time, Markov, Stochastic Systems," J. Electronics Control, (10): 473–88 (1961).
31. A. Charnes, W. W. Cooper, and G. L. Thompson, "Constrained Generalized Medians and Linear Programming under Uncertainty," Mgt. Sci., in press.
32. ——, and M. Kirby, "Optimal Decison Rules for the E Model of Chance-Constrained Programming," Research Analysis Corporation, RAC-TP-166, Jul 65.
33. S. Perlis, Theory of Matrices, Addison-Wesley Press, Inc., Cambridge, Mass., 1958.
34. R. M. Thrall and L. Tornheim, Vector Spaces and Matrices, John Wiley & Sons, Inc., New York, 1957.
35. A. S. Householder, Principles of Numerical Analysis, McGraw-Hill Book Company, New York, 1953.
36. K. Hoffman and R. Kunze, Linear Algebra, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961.
37. A. Charnes and W. W. Cooper, Management Models and Industrial Applications of Linear Programming, John Wiley & Sons, Inc., New York, 1961.

ADDITIONAL REFERENCES

Ben-Israel, A., and Y. Ijiri, "A Report on the Machine Computation of the Generalized Inverse of an Arbitrary Matrix," ONR Research Memo. 110, Carnegie Institute of Technology, Pittsburgh, Penna. (1963).

Charnes, A., and M. Kirby, "A Linear Programming Application of a Left Inverse of a Bases Matrix," ONR Research Memo. 91, Northwestern University, the Technological Institute, Evanston, Ill. (1963).

——, ——, "The Dual Method and the Method of Balas and Ivanescu for the Transportation Model," Cahiers du centre d'Etudes de Recherche Operationelle, 6 (1): 5–18 (1964).

——, ——, "Modular Design, Generalized Inverses and Convex Programming," Research Analysis Corporation, RAC-TP-157, in press.

Hadley, R., Linear Algebra, Addison-Wesley Press, Inc., Cambridge, Mass., 1961.