1-1977

GENERALIZED INVERSES OF MATRICES AND THEIR APPLICATIONS

Eleanor L. White
Longwood University

GENERALIZED INVERSES OF MATRICES AND THEIR APPLICATIONS

By

Eleanor L. White

January 1977

A paper presented to Longwood College in partial fulfillment of the requirements for honors in mathematics

ACKNOWLEDGMENTS

The author would like to express her deep appreciation for her advisor, Dr. Robert P. Webber. Without his gentle encouragement, infinite patience, criticisms and suggestions this paper would never have been undertaken. THANKS!

She also wishes to thank her committee members for their encouragement and help.

Finally, she wishes to remember her friends and family who helped push while Dr. Webber pulled.

TABLE OF CONTENTS

INTRODUCTION

The generalized inverses of matrices discussed in this paper consist only of the Moore-Penrose generalized inverse defined by equations (3) - (6) and several of the more applicable inverses defined by taking one or two of these equations separately. In particular, the minimum norm and least squares generalized inverses use the first and fourth and first and third equations respectively.

Chapter 1 defines the notation which will be used throughout the paper.

Chapter 2 lists the definitions and theorems needed from linear algebra. The references for the proofs of the theorems cited are given where the proofs themselves are not given.

Chapter 3 begins the core of the paper with the definitions of a generalized inverse associated with E. H. Moore and R. Penrose. The equivalence of these definitions is shown. The uniqueness of this generalized inverse is proven and a computational method for finding the matrix representing this inverse is developed. Then after proving several useful properties of the Moore-Penrose generalized inverse, two examples are worked. The second example shows how this inverse can be used to find the best approximation of a solution of a system of inconsistent equations.

Chapters 4 and 5 discuss the minimum norm inverse and the least squares inverse respectively.

Chapter 6 contains an application of a generalized inverse as it pertains to interval programming problems. Used to solve these problems is the generalized inverse which satisfies the first equation of the Moore-Penrose generalized inverse (equation 3).

The final application, to Markov chains, is discussed in Chapter 7. The example worked at the end of the chapter uses the Moore-Penrose generalized inverse. However, a theorem in the chapter allows for the use of any generalized inverse satisfying at least one of the four Moore-Penrose equations if certain conditions are true.

Chapter 8 contains the figures referred to in chapters 4 and 5.

CHAPTER I

NOTATION

$\mathbb{C}$	field of complex numbers
$\mathbb{C}_n$	space of n-dimensional column vectors over the complex numbers
$\mathbb{C}_{mxn}$	space of m x n matrices over $\mathbb{C}$
R	field of real numbers
R_n	space of n-dimensional column vectors over the real numbers
R_{mxn}	space of m x n matrices over R
A^T	transpose of a matrix A
A^*	conjugate transpose of a matrix A
$\bar{a}$	complex conjugate of a vector a
$a^{\perp}$	orthogonal complement of a vector a
$V^{\perp}$	orthogonal complement of a subspace V
$P_X^{\perp}$	orthogonal projection of $\mathbb{C}_n$ onto a subspace X
e_j	jth column of the identity matrix
e	column vector having ones in each component
I	identity matrix
R(A)	range of A
N(A)	null space of A
A^+	generalized (Moore-Penrose) inverse
$\langle x,y \rangle$	denotes the inner product of two vectors, defined by $\langle x,y \rangle = \Sigma \overline{y_i} x_i = y^* x$ where $x = (x_i)$ and $y = (y_i)$ for $i = 1, 2, \ldots, n$.

$||x||$ denotes the Euclidean norm of a vector, defined by

$$||x|| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

where $x = (x_i)$ for $i = 1, 2, \ldots, n$.

■ denotes the end of a proof

Lower case Latin and Greek letters denote column vectors and scalars.

Upper case Latin letters denote matrices and subspaces.

All matrices and vectors will be taken over the complex numbers unless otherwise specified.

CHAPTER II

PRELIMINARIES

Within the body of the paper several definitions and theorems are needed from linear algebra. They are listed below with references given for the proofs of the theorems if the proof does not follow the stated theorem.

DEFINITION 1: The <u>inner product</u> (or scalar product) of two n x 1 column vectors u, v with real or complex elements is the scalar quantity $\langle u,v \rangle$ defined by

$$\langle u,v \rangle = \overline{u_1}v_1 + \overline{u_2}v_2 + \cdots + \overline{u_n}v_n ,$$

where $\overline{u_i}$ is the complex conjugate of u_i. The <u>norm</u> or length of u, denoted by $\|u\|$, is defined by

$$\|u\| = \langle u,u \rangle^{1/2} = (\overline{u_1}u_1 + \overline{u_2}u_2 + \cdots + \overline{u_n}u_n)^{1/2}.$$

Two vectors u and v are <u>orthogonal</u> if and only if

$$\langle u,v \rangle = 0.$$

DEFINITION 2: Let V be a vector space. W is any subset of V. The <u>orthogonal complement</u> of W, denoted by $W^{\perp}$ (read "W perp"), consists of those vectors in V which are orthogonal to every $w \in W$:

$$W^{\perp} = \{v \in V: \langle v,w \rangle = 0 \text{ for every } w \in W\}.$$

DEFINITION 3: A matrix A is <u>Hermitian</u> if and only if $A = A^*$.

THEOREM 1: Let X and Y be two matrices whose product XY is defined. Then $\text{rank}(XY) \leq \min(\text{rank}(X), \text{rank}(Y))$ [6, p. 112].

DEFINITION 4: Let V and W be vector spaces. A function $L: V \rightarrow W$ is called a linear transformation of V into W if it satisfies

1. $L(\alpha + \beta) = L(\alpha) + L(\beta)$ for α and β in V.
2. $L(c\alpha) = cL(\alpha)$ for α in V, and c a real number.

THEOREM 2: There is a one - to - one correspondence between the set of linear transformations and the set of matrices. Hence every linear transformation can be uniquely represented by a matrix [3, p. 205].

THEOREM 3: Let W be a subspace of V. Then any vector in the vector space can be uniquely represented as the sum of a vector in the subspace and a vector orthogonal to every vector in the subspace. In particular, for any $v \in V$, v can be written uniquely as

$$v = w + w^{\perp} \quad \text{for } w \in W \text{ and } w^{\perp} \in W^{\perp}$$ [6, p. 295].

DEFINITION 5: Consider a linear transformation $E: V \rightarrow W$ given by $E(v) = w$. E is the orthogonal projection of V onto W provided

1. if $w \in W$, then $E(w) = w$; and
2. if $z \in W^{\perp}$, then $E(z) = 0$.

THEOREM 4: Given a space V and a subspace W then there is a unique orthogonal projection of V onto W.

Proof: Let $v \in V$. By Theorem 3 there exists a unique w and $w^{\perp}$ in W and $W^{\perp}$ respectively such that $v = w + w^{\perp}$. Let E be a linear transformation of V into W such that $E(v) = w$. Then E satisfies the conditions of Definition 5 and thus E is an orthogonal projection onto W. Suppose F is also an orthogonal projection onto W. Then $F(v) = F(w + w^{\perp}) = w = E(v)$. So $F = E$ for every v. ∎

THEOREM 5: A projection is a linear transformation and hence can be uniquely expressed as a matrix [Definition 5 and Theorem 2].

DEFINITION 6: A <u>major determinant</u> of a p x q matrix is the determinant of any square submatrix of maximum order.

DEFINITION 7: If A is an m x n matrix, and B is an n x m matrix with $m \leq n$, then a major determinant of A and a major determinant of B are said to be <u>corresponding majors</u> of A and B if and only if the columns of A used to form the major of A have the same indices as do the rows of B used to form the major of B.

THEOREM 6: If A is an m x n matrix and B is an n x m matrix, and if $m \leq n$, then the determinant of AB is equal to the sum of the products of the corresponding majors of A and B [3, pp. 64 - 65].

THEOREM 7: If $L: V \to W$ is a linear transformation of an n-dimensional vector space V into a vector space W, then the dimension of the null space of L plus the dimension of the domain of L equals the dimension of V [4, p. 100].

THEOREM 8: If P is nonsingular, then A and PA have the same column rank [9, p. 55].

THEOREM 9: If L is a linear transformation of a vector space V over ¢ such that $\langle Lv, v \rangle = 0$ for all $v \in V$, then $L = 0$. [12, p. 167].

LEMMA 1: If x and y are two orthogonal vectors, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Proof:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle ; \\ &= \langle x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y \rangle ; \\ &= \|x\|^2 + \|y\|^2 . \end{aligned}$$ ■

CHAPTER III

THE MOORE-PENROSE GENERALIZED INVERSE

In 1920 E. H. Moore published a paper on a generalized inverse for matrices [7] . This inverse could be found for any singular and rectangular matrix. Until this time only inverses of nonsingular matrices had been studied. The following definition of a generalized inverse is due to Moore.

DEFINITION (Moore): A matrix X is a generalized inverse of A if

$$XA = P^{\perp}_{R(X)} \tag{1}$$

and

$$AX = P^{\perp}_{R(A)}. \tag{2}$$

Several years later in 1955, R. Penrose, unaware of Moore's work, developed the same generalized inverse [8] . However he defined it in a slightly different manner.

DEFINITION (Penrose): A matrix X is a generalized inverse of A if

$$AXA = A, \tag{3}$$

$$XAX = X, \tag{4}$$

$$(AX)^* = AX, \tag{5}$$

and

$$(XA)^* = XA. \tag{6}$$

Because of the independence of the work, this inverse is often referred to as the Moore-Penrose inverse.

The claim that Moore's definition is equivalent to Penrose's definition of a generalized inverse is by no means obvious. Before establishing the equivalence of these two definitions several theorems must be proven.

THEOREM 10: For any matrix $A \in \mathbb{C}_{nxn}$, $N(A^*) = R(A)^{\perp}$.

Proof: By definition of $N(A^*)$ a vector $x \in \mathbb{C}_n$ is in $N(A^*)$ if and only if $A^*x = 0$. This is equivalent to the statement that $x \in N(A^*)$ if and only if each row of A^* postmultiplied by x gives the product 0. Now, the rows of A^* are the conjugate transposes of the columns of A, and therefore $x \in N(A^*)$ if and only if it is orthogonal to every column of A. Thus $x \in N(A^*)$ if and only if $x \in R(A)^{\perp}$. Hence $N(A^*) = R(A)^{\perp}$. ∎

THEOREM 11: If a matrix is idempotent and Hermitian then it is an orthogonal projection onto its range.

Proof: Let X be idempotent and Hermitian. Then to be an orthogonal projection, X must satisfy the following conditions:

1. if $x \in R(X)$, then $Xx = x$

and

2. if $z \in R(X)^{\perp}$, then $Xz = 0$.

Let $x \in R(X)$. Then there exists a $y \in \mathbb{C}_n$ such that $x = Xy$. But X is idempotent so $x = Xy = X(Xy) = Xx$. So X acts as an identity on its range.

Let $z \in R(X)^{\perp}$. But $R(X)^{\perp} = N(X^*)$ by Theorem 10. Since X is Hermitian $N(X^*) = N(X)$, so that $z \in N(X)$. Thus $Xz = 0$. ∎

THEOREM 12: An orthogonal projection P is idempotent and Hermitian.

Proof: Let P be an orthogonal projection of V onto W. Define the jth column of P to be p_j. Since $Pe_j = p_j$, $p_j \in R(P)$. Thus

$$Pp_j = p_j \quad \text{for each } j$$

by definition of an orthogonal projection. Written in matrix notation the equation becomes $PP = P$ or $P^2 = P$. Hence P is idempotent.

If $v \in V$, then v can be expressed uniquely as

$$v = w + w^{\perp}$$

where $w \in R(W)$ and $w^{\perp} \in R(W)^{\perp}$. Hence

$$\begin{aligned} \langle P^*v, v\rangle = \langle v, Pv\rangle &= \langle v, w\rangle = \langle w + w^{\perp}, w\rangle; \\ &= \langle w, w\rangle + \langle w^{\perp}, w\rangle; \\ &= \langle w, w\rangle = \langle w, w\rangle + \langle w, w^{\perp}\rangle; \\ &= \langle w, w + w^{\perp}\rangle = \langle w, v\rangle; \\ &= \langle Pv, v\rangle, \end{aligned}$$

giving

$$\langle P^*v, v\rangle = \langle Pv, v\rangle,$$

and hence

$$\langle (P^* - P)v, v\rangle = 0,$$

for all $v \in V$. By Theorem 9,

$$(P^* - P) = 0$$

so $P^* = P$. Hence P is Hermitian. ∎

THEOREM 13: $MN = P^{\perp}_{R(M)}$ if and only if $MNM = M$ and $(MN)^* = MN$.

Proof: i. MN is an orthogonal projection onto $R(M)$. Let m_j be the jth column of M so $m_j \in R(M)$. Thus $MNm_j = m_j$. Written in matrix notation the equation becomes $MNM = M$. By Theorem 12, $(MN)^* = MN$.

ii. Since $(MN)^* = MN$, MN is Hermitian. Multiplying both sides of $MNM = M$ on the right by N yields $MNMN = MN$. Thus $(MN)^2 = MN$. Therefore MN is idempotent. Hence MN is an orthogonal projection onto $R(MN)$.

Let $x \in R(MN)$. Then $x = MNy$ for some y, so $x = M(Ny) \in R(M)$. Hence $R(MN) \subseteq R(M)$. Conversely, let $a \in R(M)$. Then $a = Mb$. for some b, so

$$a = Mb = MNMb = MN(Mb).$$

Hence $a \in R(MN)$ and $R(M) \subseteq R(MN)$. Therefore $R(M) = R(MN)$.

Thus, MN is an orthogonal projection onto $R(M)$. Since there is a unique projection onto $R(M)$, $MN = P^{\perp}_{R(M)}$. ∎

By applying Theorem 13 in the following corollaries the two definitions are shown to be equivalent.

COROLLARY 1: $AX = P^{\perp}_{R(A)}$ if and only if $AXA = A$ and $(AX)^* = AX$.

COROLLARY 2: $XA = P^{\perp}_{R(X)}$ if and only if $XAX = X$ and $(XA)^* = XA$.

The equivalence of the two definitions of the generalized inverse has been shown in Corollaries 1 and 2. The rest of this chapter consists of a method for finding the orthogonal projections onto either the range of X or the range of A, a theorem which shows the uniqueness of the Moore-Penrose generalized inverse, and several useful properties of the Moore-Penrose generalized inverse. Then several examples dealing with the computation of this generalized inverse and its use in solving inconsistent systems of linear equations are presented.

For any m x n matrix A over $\mathbb{C}$ a representation of the orthogonal projection onto the range of A may be obtained using the following method.

Let A be an m x n matrix over $\mathbb{C}$ with rank r. The r linearly independent column vectors which form a basis of the column space or range R(A) of A also form a matrix, $B = [b_1 \ b_2 \ . \ . \ . \ b_r]$, having rank r. Every column vector of A may be written as a linear combination of the columns of B; i.e., $a_j = Bc_j$ where c_j, the jth column of a r x n matrix C, is determined by expressing a_j as a linear combination of the basis vectors in B. Therefore A = BC.

Using Theorem 1, the rank of C is found to be r. Indeed,

$$\begin{aligned} r &= \text{rank of } A \quad \text{by definition of } A; \\ &= \text{rank of } BC; \\ &\leq \text{rank of } C \quad \text{by Theorem 1:} \\ &\leq r \quad \text{since } C \text{ is an } r \times n \text{ matrix.} \end{aligned}$$

Therefore, $r \leq$ rank of $C \leq r$, so the rank of C is r.

Taking the conjugate transpose of A = BC, the equation

$A^* = C^*B^*$ is obtained and the columns of C^* form a basis for the range of A^*, written $R(A^*)$. (The jth column of A^* is written as a linear combination of the vectors of C^* with the jth column of B^*; $a_j^* = C^*b_j^*$.)

In general a column vector of $R(A)$ is of the form $a = Bx$ for $x \in \mathbb{C}_n$. Therefore, the problem may be restated: For any $s \in \mathbb{C}_n$, find the vector $Bx \in R(A)$ which is a projection of s in $R(A)$. In particular, s may be written uniquely as the sum of a vector in $R(A)$ and a vector in $R(A)^{\perp}$,

$$s = a + a^{\perp} \quad \text{where } a \in R(A) \text{ and } a^{\perp} \in R(A)^{\perp} .$$

But $a = Bx$ for some x. So $s = Bx + a^{\perp}$. Subtracting Bx from both sides,

$$s - Bx = a^{\perp} .$$

Hence $s - Bx$ is orthogonal to every vector in $R(A)$ and therefore to each of the basis vectors. Thus Bx may be determined as follows:

$$\langle b_j, s - Bx \rangle = 0, \text{ for } j = 1, 2, \ldots, r.$$

By applying the definition of inner product, the equation becomes

$$b_j^*(s - Bx) = 0$$

which can be written as a matrix equation

$$B^*(s - Bx) = 0.$$

Using the distributive property of matrices and adding B^*Bx to both sides of the equation,

$$B^*s = B^*Bx. \tag{7}$$

Since B^*B is an r-square matrix its rank cannot exceed r.

Assume the rank of B*B to be k where k r. Then by Theorem 6 the determinant of B*B is the sum of the products of the corresponding majors of B* and B. Since in this case pairs of corresponding majors are complex conjugate numbers, their products are nonnegative. Since B is of full column rank, some major M of B is nonsingular. Thus M* is nonsingular, and the determinant of M* times the determinant of M is greater than 0. Thus the determinant of B*B is greater than 0. Therefore B*B is nonsingular and hence invertible. Therefore, equation (7) may be rewritten

$$(B^*B)^{-1}B^*s = x.$$

Using Theorem 5, P is defined as an n x n matrix which projects s onto R(A). Thus

$$Ps = Bx = B(B^*B)^{-1}B^*s$$

for all s. Therefore for any column vector s,

$$Ps = B(B^*B)^{-1}B^*s.$$

Let $s = e_j$. Then $Pe_j = B(B^*B)^{-1}B^*e_j$ implies that the jth column of P equals the jth column of $B(B^*B)^{-1}B^*$ for j = 1, 2, . . . , r. Therefore,

$$P = B(B^*B)^{-1}B^*.$$

The matrix P is an orthogonal projection. Indeed,

$$P^2 = (B(B^*B)^{-1}B^*)^2 = (B(B^*B)^{-1}B^*)(B(B^*B)^{-1}B^*);$$
$$= B(B^*B)^{-1}(B^*B)(B^*B)^{-1}B^* = B(B^*B)^{-1}B^* = P$$

and

$$P^* = (B(B^*B)^{-1}B^*)^* = B(B^*B)^{-1}B^* = P$$

shows that P is idempotent and Hermitian.

By a similar method a projection onto $R(A^*)$ can be found. Call such a projection Q. As above,

$$Q = C^*(CC^*)^{-1}C$$

and C^* is the matrix containing a basis of $R(A^*)$ as its columns.

For any matrix A it is shown in the development of an orthogonal projection onto $R(A)$ that any vector in $R(A)$ can be written as a linear combination of the basis vectors of $R(A)$. Thus $A = BC$, where B is a matrix containing the basis vectors of $R(A)$ as its columns. Having defined the matrices B and C it is possible to define a matrix A^+ such that A^+ is a generalized inverse of A.

THEOREM 14: Let A be an $m \times n$ matrix, P be the projection onto $R(A)$, and Q be the projection onto $R(A^*)$. Then there exists a unique matrix A^+ such that

i. $AA^+A = A$,

ii. $A^+AA^+ = A^+$,

iii. $AA^+ = P$, and

iv. $A^+A = Q$.

Proof: Let $A^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$. It can be shown that this matrix satisfies the four properties. Indeed,

i. $AA^+A = BC\,C^*(CC^*)^{-1}(B^*B)^{-1}B^*\,BC$;

$$= B(CC^*)(CC^*)^{-1}(B^*B)^{-1}(B^*B)C = BC = A.$$

ii. $A^+AA^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^* \; BC \; C^*(CC^*)^{-1}(B^*B)^{-1}B^*$;

$= C^*(CC^*)^{-1}(B^*B)^{-1}(B^*B)(CC^*)(CC^*)^{-1}(B^*B)^{-1}B^*$;

$= C^*(CC^*)^{-1}(B^*B)^{-1}B^* = A^+$.

iii. Since $A = BC$, $AA^+ = BC \; C^*(CC^*)^{-1}(B^*B)^{-1}B^*$;

$= B(B^*B)^{-1}B^* = P$.

iv. Since $A = BC$, $A^+A = C^*(CC^*)^{-1}(B^*B)^{-1}B^* \; BC$;

$= C^*(CC^*)^{-1}C = Q$.

Having shown that A^+ satisfies the four properties listed above, assume two such matrices, A_1^+ and A_2^+, exist and that both satisfy the four properties. Since $AA_1^+ = P$,

$$A_2^+AA_1^+ = A_2^+P.$$

Moreover, $AA_2^+ = P$, and $A_2^+AA_2^+ = A_2^+P$. Therefore,

$$A_2^+AA_1^+ = A_2^+AA_2^+ = A_2^+.$$

But $A_2^+A = Q$, and hence $A_2^+AA_1^+ = QA_1^+$. Similarly $A_1^+A = Q$, and hence $A_1^+AA_1^+ = QA_1^+$. This implies $QA_1^+ = A_1^+$ so that $A_2^+AA_1^+ = A_1^+$. However, it has been shown that $A_2^+AA_1^+ = A_2^+$. Therefore, $A_1^+ = A_2^+$. Thus A^+ is unique. Therefore A^+ is the unique matrix that satisfies the four properties. ∎

In linear algebra an inverse A^{-1} is defined only for a nonsingular matrix A, and A^{-1} is unique. However a matrix also has a generalized inverse A^+ which is also unique. The following theorem will resolve this seeming contradiction of two unique inverses for the same nonsingular matrix.

THEOREM 15: If a matrix A is nonsingular, then $A^+ = A^{-1}$.

Proof: If A^{-1} exists, then

i. $AA^{-1}A = AI = A$.

ii. $A^{-1}AA^{-1} = A^{-1}I = A^{-1}$.

iii. $AA^{-1} = I = P$ since for $x \in R(A)$, $Px = Ix = x$.

iv. $A^{-1}A = I = Q$ since for $y \in R(A^*)$, $Qy = Iy = y$. Thus A^{-1} satisfies the four equations in Theorem 14. Since A^+ is the unique matrix satisfying these equations, then $A^{-1} = A^+$. ■

Even though A^{-1} is defined only for nonsingular matrices; A^+ may be determined for singular matrices. Being able to determine A^+ is useful when finding solutions of a system of linear equations. There are three possibilities for the number of solutions a system of linear equations may have. If $Ax = b$ is consistent, then either one solution exists or an infinite number exist. In either case for any x which is a solution of the system, $\| Ax - b \| = 0$. If there are an infinite number of solutions then among them there exists a x such that $\| x \|$ is a minimum. However, if $Ax = b$ is inconsistent, then no solution exists. It will be shown that A^+b minimizes $||Ax - b||$. Among the vectors which minimize $||Ax - b||$ $x = A^+b$ is the unique vector with minimum norm.

LEMMA 2: Let $A \in \mathbb{C}_{mxn}$ and P, Q, and A^+ be defined as in Theorem 14. Then

i. $R(I - P) = R(A)^{\perp}$,

ii. $(I - P)$ is a projection onto $R(A)^{\perp}$,

iii. $R(A^+) = R(A^*)$, and

iv. $(I - Q)$ is a projection onto $R(A^+)^{\perp}$.

Proof: i. P is a projection onto R(A). R(A) is a subspace, so by Theorem 3, a vector $s \in \mathbb{C}_n$ can be uniquely expressed as the sum of a vector in the subspace and a vector in its orthogonal complement. So $s = x + x^{\perp}$. But x can be expressed as the projection of s onto R(A) or $Ps = x$. Substituting Ps for x the equation becomes $s = Ps + x^{\perp}$. Subtracting Ps from both sides, $s - Ps = x^{\perp}$. Using the distributive property of matrices $(I - P)s = x^{\perp}$. So $R(I - P) \subseteq R(A)^{\perp}$.

Let $x \in R(A)^{\perp}$, then $Px = 0$. So $(I - P)x = x$. Hence $x \in R(I - P)$. So $R(A)^{\perp} \subseteq R(I - P)$. Therefore $R(A)^{\perp} = R(I - P)$.

ii. Since

$$(I - P)^2 = (I - P)(I - P) = I - 2P + P^2 ;$$
$$= I - 2P + P = (I - P)$$

and

$$(I - P)^* = (I^* - P^*) = (I - P)$$

$(I - P)$ is idempotent and Hermitian. Therefore $(I - P)$ is an orthogonal projection onto the orthogonal complement of $R(A)$.

iii. By definition Q is a projection onto $R(A^*)$ and $Q = A^+A$; but Q is also a projection onto $R(A^+)$. Indeed for any $y \in R(A^*)$,

$$y = Qy = A^+Ay.$$

This implies that $y \in R(A^+)$. So $R(A^*) \subseteq R(A^+)$.

However, for any $y \in R(A^+)$, $y = A^+z$ for some z, so that

$$y = A^+AA^+z = Q(A^+)z,$$

implying $y \in R(A^*)$. So $R(A^+) \subseteq R(A^*)$. Hence $R(A^*) = R(A^+)$.

iv. By a proof similar to part ii, $(I - Q)$ is a projection onto $R(A^+)^{\perp}$. ∎

THEOREM 16: Let $x_o = A^+b$. Then for any $x \neq x_o$, either

i. $\| Ax - b \| > \| Ax_o - b \|$, or

ii. $\| Ax - b \| = \| Ax_o - b \|$ and $\| x \| > \| x_o \|$.

Proof: i. For any x,

$$Ax - b = A(x - A^+b) + (I - AA^+)(-b).$$

Making the substitution $x_o = A^+b$,

$$Ax - b = A(x - x_o) + (I - AA^+)(-b).$$

But $AA^+ = P$, so

$$Ax - b = A(x - x_o) + (I - P)(-b).$$

Hence

$$\| Ax - b \|^2 = \| A(x - x_o) + (I - P)(-b) \|^2.$$

By part i of Lemma 2, $A(x - x_o)$ is orthogonal to $(I - P)(-b)$.

Applying Lemma 1, the equation becomes

$$\|Ax - b\|^2 = \|A(x - x_o)\|^2 + \|(I - P)(-b)\|^2 ;$$
$$= \|A(x - x_o)\|^2 + \|(Pb - b)\|^2 ;$$
$$= \|A(x - x_o)\|^2 + \|AA^+b - b\|^2 ;$$
$$= \|A(x - x_o)\|^2 + \|Ax_o - b\|^2 .$$

Hence $\|Ax - b\| > \|Ax_o - b\|$ unless $Ax = Ax_o$.

ii. If $Ax = Ax_o$, then

$$A^+Ax = A^+Ax_o .$$

By substitution $A^+Ax_o = A^+AA^+b$. But $A^+AA^+ = A^+$, so

$$A^+Ax = A^+Ax_o = A^+b = x_o .$$

The vector x may be written

$$x = x_o + (x - x_o) = A^+b + (x - A^+Ax),$$

so $x = A^+b + (I - A^+A)x$. But $A^+A = Q$. Using this substitution,

$$x = A^+b + (I - Q)x.$$

Squaring the norm of both sides,

$$\|x\|^2 = \|A^+b + (I - Q)x\|^2 .$$

By part iii of Lemma 2, A^+b is orthogonal to $(I - Q)x$.

Applying Lemma 1, the equation becomes

$$\|x\|^2 = \|A^+b\|^2 + \|(I - Q)x\|^2 ;$$
$$= \|x_o\|^2 + \|x - Qx\|^2 ;$$
$$= \|x_o\|^2 + \|x - A^+Ax\|^2 ;$$

$$= ||x_o||^2 + ||x - A^+b||^2 ;$$

$$= ||x_o||^2 + ||x - x_o||^2 .$$

Therefore $||x|| > ||x_o||$ if $x \neq x_o$. Thus among the vectors which minimize $||Ax - b||$ x_o is the unique vector which has minimum norm.

EXAMPLE:

Find the generalized inverse of

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

Solution: Upon column reducing A it is found that

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

The matrix C is obtained by finding the coefficients needed to write A as a linear combination of B. The matrix C that was found is

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

Applying the definition of the Moore-Penrose generalized inverse,

$$A^{+} = C^{*}(CC^{*})^{-1}(B^{*}B)^{-1}B^{*},$$

it is found that

$$A^{+} = \frac{1}{102}\begin{bmatrix} -15 & -18 & 3 & -3 & 18 & 15 \\ 8 & 13 & -5 & 5 & -13 & -8 \\ 7 & 5 & 2 & -2 & -5 & -7 \\ 6 & -3 & 9 & -9 & 3 & -6 \end{bmatrix}.$$

The general form of a linear equation can be written

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n - b = 0.$$

A system of linear equations would contain m such linear equations. In matrix notation the system of linear equations would be written

$$Ax - b = 0$$

where A is the mxn matrix of coefficients, x is a nx1 matrix of variables, and b is a mx1 matrix of constants.

If a system of linear equations, $Ax = b$, has no solution, then any nx1 vector x will cause Ax to differ from b by some residual amount, r, where $r = Ax - b$. When r is "small" this vector, x, can be thought of as an approximate solution. Since an infinite number of such solutions, could exist, it is helpful to look at a subset of all the approximate solutions. Let that subset be the set of all vectors which minimize $\|Ax - b\|$. By minimizing $\|Ax - b\|$, the vector x minimizes the sum of the squares of the components of the residual vector r.

In order to attach meaning to the residual vector, consider that its jth component has the form

$$a_{j1}x_1 + a_{j2}x_2 + \cdot \cdot \cdot + a_{jn}x_n - b_j$$

where $x_1, x_2, \cdot \cdot \cdot , x_n$ are components of the vector x, $a_{j1}, a_{j2}, \cdot \cdot \cdot , a_{jn}$ are the coefficients of the jth equation, and b_j is the jth component of the vector b. In n-space geometry the distance between the point with coordinates $(\alpha_1, \alpha_2, \cdot \cdot \cdot , \alpha_n)$ and the n-space representation of the equation $a_1x_1 + a_2x_2 + \cdot \cdot \cdot + a_nx_n = b$ is given by the formula

$$d = \frac{|a_1\alpha_1 + a_2\alpha_2 + \cdot \cdot \cdot + a_n\alpha_n - b|}{\sqrt{a_1^2 + a_2^2 + \cdot \cdot \cdot + a_n^2}} .$$

Furthermore, if the equation is normalized, i.e.,

$$\sqrt{a_1^2 + a_2^2 + \cdot \cdot \cdot + a_n^2} = 1,$$

the formula reduces to

$$d = |a_1\alpha_1 + a_2\alpha_2 + \cdot \cdot \cdot + a_n\alpha_n - b| .$$

Therefore, if the equations of the system $Ax - b = 0$ are normalized, the absolute value of the jth component of the residual vector represents the distance from the point defined by the vector x and the n-space representation of the jth equation.

To rewrite a linear equation in its normal form, find the square root of the sum of the squares of the coefficients of the variables, i.e., $k = \pm\sqrt{a_1^2 + a_2^2 + \cdot \cdot \cdot + a_n^2}$. The normal form can then be written by dividing each term in the equation by the k which has the opposite sign from the constant of the equation. Now that the normalized equations have been found, substitute the corresponding components of the vector x, into each equation to find the corresponding component of r.

After writing each equation in a system of equations in its normal form, rewrite the normalized system of equations in matrix form, $Ax = b$. The set of all vectors which minimize $\|Ax - b\|$ could be an infinite set. However, if the Moore-Penrose generalized inverse is used to solve for x, the unique vector of minimum norm obtained will be in the set consisting of all vectors which minimize $\|Ax - b\|$.

For any solution vector, x, of a system of linear equations to have physical meaning, i.e., for the residual vector to represent the undirected distances from a point defined by a vector x to the n-space representation of the jth linear equation, the normalized equations are used when solving for x. Thus in working examples, the normalized form of the equations will be found before the Moore-Penrose inverse is obtained. The Moore-Penrose generalized inverse will be used to find the unique vector of minimum norm in the set of vectors which minimize $\|Ax - b\|$.

EXAMPLE:

Find the unique solution vector of minimum norm that minimizes $\|Ax - b\|$ for the following system of equations:

$$3x + 2y = 6$$
$$4x - 8y = 8$$
$$x = 0.$$

Solution: Normalizing each equation yields

$$\frac{3}{\sqrt{13}}x + \frac{2}{\sqrt{13}}y = \frac{6}{\sqrt{13}}$$

$$\frac{1}{\sqrt{5}}x - \frac{2}{\sqrt{5}}y = \frac{2}{\sqrt{5}}$$

$$x = 0.$$

The normalized system of equations can be rewritten in matrix form as

$$\begin{bmatrix} \frac{3}{\sqrt{13}} & \frac{2}{\sqrt{13}} \\ \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{6}{\sqrt{13}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}.$$

Factoring A into BC it is found that

$$B = \begin{bmatrix} \frac{3}{\sqrt{13}} & \frac{2}{\sqrt{13}} \\ \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Upon substituting into the equation $A^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$,

$$A^+ = \begin{bmatrix} \frac{3}{\sqrt{13}} & \frac{1}{\sqrt{5}} & 1 \\ \frac{2}{\sqrt{13}} & \frac{-2}{\sqrt{5}} & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{13}} & \frac{2}{\sqrt{13}} \\ \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \frac{3}{\sqrt{13}} & \frac{1}{\sqrt{5}} & 1 \\ \frac{2}{\sqrt{13}} & \frac{-2}{\sqrt{5}} & 0 \end{bmatrix};$$

$$= \begin{bmatrix} .42418 & .26307 & .52941 \\ .47721 & -.82208 & -.02941 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{+}b = \begin{bmatrix} .42418 & .26307 & .52941 \\ .47721 & -.82208 & -.02941 \end{bmatrix} \begin{bmatrix} \frac{6}{\sqrt{13}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}.$$

Therefore, $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} .94118 \\ .05883 \end{bmatrix}$ is the solution vector of minimum norm which minimizes $\|Ax - b\|$. Figure 10 is a graph of these equations with the best approximate solution marked.

CHAPTER IV

THE MINIMUM NORM INVERSE

A minimum norm inverse of a matrix A is defined as a matrix Y such that A = AYA and (YA)* = YA. Therefore the minimum norm inverse satisfies the first and fourth properties of the Moore-Penrose inverse, equations (3) and (6). Hence the minimum norm inverse shall be denoted by $A^{(1,4)}$.

Norms may be thought of as distances. Thus a minimum norm is the smallest distance. When working with solutions of a system of linear equations the minimum norm solution represents the solution which when expressed as a vector has the smallest length. Here norm is defined by

$$\|x\| = \langle x,x\rangle^{1/2}$$

where the inner product is defined by

$$\langle x,y\rangle = \sum \overline{y_i} x_i = y^* x.$$

If a system of linear equations Ax = b has a nonempty solution set, then by definition it is consistent. If there is only one solution in the set, then that solution is trivially the minimum norm solution (figure 1). An example is the intersection of two nonparallel lines. However, if several solutions exist for a system of linear equations one of the solutions must be the minimum norm solution. This becomes clearer if the graph of coinciding lines in 2-space is considered (figure 2). An infinite number of points are in the solution set, yet only one point is closest to the origin and hence the term minimum norm.

These two examples have dealt with systems of linear equations in 2-space. A linear equation represents a line in 2-space and a system of linear equations represents a family of lines. If the concept of space is extended to 3-space a linear equation represents a plane and therefore a system of linear equations represents a family of planes.

Figures 3 and 4 show the various ways three planes may intersect and a solution or solutions exist. In figure 3 the three planes have a unique point of intersection which causes the minimum norm solution to be trivial. However, two cases exist where the solution set contains an infinite number of points and hence the minimum norm solution is not trivial. Consider the case in which all three planes intersect in a line (figure 4). Since the intersection is a line, one of the points on the line must be closest to the origin. This point will be the solution of minimum norm. The other case to be considered deals with the intersection of coinciding planes. The intersection is a plane and the solution set consists of all points in the plane. There must be one point in the plane closest to the origin - the minimum norm solution.

Any consistent system of linear equations will have a minimum norm solution. However, not all systems are consistent. If a system of linear equations is inconsistent, then no solution exists and therefore no minimum norm solution exists. Examples of systems of equations which have no minimum norm solution can easily be found in both 2- and 3- space. In two dimensions two parallel lines do not intersect (figure 5). Therefore the system of equations which represents this family of lines has no solution. Another example in 2- space is found by extending the line segments which form the sides of a polygon (figures 6a and 6b). There is no common point of intersection. Therefore the system of equations which represents this family of lines has no solution. (This example is mentioned since the system of equations contains more equations than unknowns. Thus the coefficient matrix is nonsquare.) Since no solution exists in each of these cases in 2- space, no minimum norm solution exists.

Similarly in 3- space parallel planes do not intersect. Even if two parallel planes are cut by a third plane there is no common intersection (figures 7 and 8). Upon considering the figure formed by n planes which intersect in n mutually parallel lines (figure 9), it is found that no common intersection is found. If the system of equations which represents the planes in any of the above cases is examined, it is found that no solution exists, and therefore no minimum norm solution exists.

It is important that a distinction be made between a minimum norm inverse and a minimum norm solution. A minimum norm inverse exists for any matrix A. However, for a system of linear equations, $Ax = b$, to have a minimum norm solution, $Ax = b$ must have a solution; i.e., be consistent. Therefore not every system of linear equations will have a minimum norm solution.

THEOREM 17: If a solution of a system of linear equations, $Ax = b$ exists, then the unique vector for which $\|x\|$ is smallest is given by

$$x_o = A^{(1,4)}b.$$

Proof: $Ax = b$ is consistent. A solution is given by $x_o = A^{(1,4)}b$ since

$$Ax_o = AA^{(1,4)}b = AA^{(1,4)}Ax = Ax = b.$$

The general solution x may be written

$$x = x_o + y$$

where $y \in N(A)$ [12, p. 107]. Then

$$\begin{aligned} \|x\|^2 &= \|x_o + y\|^2; \\ &= \langle x_o + y, x_o + y \rangle; \\ &= \langle x_o, x_o + y \rangle + \langle y, x_o + y \rangle; \\ &= \|x_o\|^2 + \|y\|^2 + \langle x_o, y \rangle + \langle y, x_o \rangle; \\ &= \|x_o\|^2 + \|y\|^2. \end{aligned}$$

Therefore $\|x\| > \|x_o\|$ unless $x = x_o$.

Given a system of consistent linear equations, $Ax = b$, it has been shown that $x_o = A^{(1,4)}b$ is the unique vector for which $\|x\|$ is a minimum. Since the minimum norm inverse is not unique, any minimum norm inverse of the matrix A when multiplied by the vector b yields x_o. Let X and Y be two such minimum norm inverses of A. Then $Xb = Yb$. Subtracting Yb from both sides gives

$$(X - Y)b = 0.$$

Thus the vector b must be in the null space of the difference of any two minimum norm inverses of A. ∎

CHAPTER V

THE LEAST SQUARES INVERSE

A least squares inverse of a matrix A is any matrix X such that $P_{R(A)}^{\perp} = AX$. From Corollary 1 this definition can be written as $AXA = A$ and $(AX)^* = AX$. Since X satisfies the first and third properties of the Moore-Penrose inverse, equations (3) and (5), a least squares inverse shall be denoted by $A^{(1,3)}$.

A least squares solution of a system of equations $Ax = b$ is a vector which causes the sum of the squares of the residual vector r to be a minimum, where $r = Ax - b$. The absolute value of the jth component of a residual vector r represents the undirected distance between a point defined by the vector, x, and the n-space representation of the jth linear equation in a system of linear equations. Thus r consists of the undirected distances between a point defined by the vector x and the n-space representations of each linear equation in a system of equations. In order to find these undirected distances, it is necessary to write each linear equation in its normal form before determining the distances. Hence the normalized equations are used to find the solution vector if the system of equations is consistent or the approximate solution vector if the system of equations is inconsistent.

If a system of linear equation, $Ax = b$, is consistent, then $r = Ax - b = 0$. Hence x is trivially a least squares solution. But if $Ax = b$ is inconsistent, then no solution exists. Then for any x, a residual vector r exists such that $r \neq 0$. Thus any x that is chosen will cause Ax to differ from b by some residual amount. The following theorem shows that the least squares inverse can be used to find a vector x which will minimize $\|Ax - b\|$.

THEOREM 18: If $Ax = b$ is a system of linear equations, then a vector which causes $\|Ax - b\|$ to be a minimum is given by $x_0 = A^{(1,3)}b$.

Proof: In general,

$$Ax - b = A(x - A^{(1,3)}b) + (I - AA^{(1,3)})(-b).$$

Making the substitution $x_0 = A^{(1,3)}b$,

$$Ax - b = A(x - x_0) + (I - AA^{(1,3)})(-b).$$

By definition $AA^{(1,3)} = P$, so

$$Ax - b = A(x - x_0) + (I - P)(-b).$$

Squaring the norm of both sides, the equation becomes

$$\|Ax - b\|^2 = \|A(x - x_0) + (I - P)(-b)\|^2.$$

By part i of Lemma 2, $A(x - x_0)$ is orthogonal to $(I - P)(-b)$. Applying Lemma 1, the equation becomes

$$\begin{aligned}\|Ax - b\|^2 &= \|A(x - x_0)\|^2 + \|(I - P)(-b)\|^2 ;\\ &= \|A(x - x_0)\|^2 + \|(Pb - b)\|^2 ;\\ &= \|A(x - x_0)\|^2 + \|AA^{(1,3)}b - b\|^2 ;\\ &= \|A(x - x_0)\|^2 + \|Ax_0 - b\|^2.\end{aligned}$$

Therefore $\|Ax - b\| > \|Ax_0 - b\|$ unless $Ax = Ax_0$. ∎

There are several ways to determine a least squares solution. Several methods are described in Ben-Israel and Greville [2, pp. 105 - 113].

Even though a least squares solution exists for all systems of equations, a least squares solution is not necessarily unique. If a system of equations is consistent, then either the solution is unique, causing the least squares solution to be unique, or there are an infinite number of solutions, causing an infinite number of least squares solutions to exist. The most interesting least squares solutions are those solutions for inconsistent, normalized systems of equations. Figures 5, 6a, and 6b show 2-space representations of systems of equations which have no solution. Similarly, figures 7 - 9 are the 3-space representations of inconsistent, normalized systems in that space. It is interesting to note that in each case there are an infinite number of least squares solutions except in the cases shown in figures 6a and 6b. These figures in 2-space represent systems of equations containing more equations than unknowns. In each case if the coefficient matrix were column reduced it would be found to have full column rank. In comparison the least squares solution of three parallel planes has an infinite number of solutions since the column reduced matrix is not of full rank.

CHAPTER VI

INTERVAL PROGRAMMING

Interval programming problems (IP) are optimization problems of the form:

$$\text{Maximize} \quad c^T x \qquad (10)$$

$$\text{subject to} \quad a \leq Ax \leq b, \qquad (11)$$

where the matrix A and column vectors a and b are given. Of particular interest is the method for solving IP problems where A has full row rank using generalized inverses.

All matrices, vectors, and scalars used in this chapter are to be taken over the field of real numbers.

DEFINITION 8: Let $F = \{ x \in R_n : a \leq Ax \leq b \}$ be the set of solutions of the IP. If F is nonempty the IP is <u>feasible</u>. Furthermore if the IP is feasible and $\max \{ c^T x : x \in F \} < \infty$ then the IP is <u>bounded</u>.

THEOREM 19: [2, p. 90]. Let $A \in R_{mxn}$, $c \in R_n$, and $a, b \in R_m$ be given such that the IP defined by (10) and (11) is feasible. Then the IP is bounded if and only if

$$c \in N(A)^{\perp} .$$

Let $A^{(1)}$ be any generalized inverse of A satisfying $AXA = A$.

THEOREM 20: [2, p. 91] . If $c \in N(A)^{\perp}$ and A is of full row rank then the optimal solutions of the IP are characterized by

$$x = A^{(1)} \eta(a, b, A^{(1)T}c) + y$$

where $y \in N(A)$ and η is defined by

$$\eta(u,v,w) = \begin{cases} u_i & \text{if} \quad w_i < 0 \\ v_i & \text{if} \quad w_i > 0 \\ \lambda_i u_i + (1-\lambda_i)v_i & \text{where } 0 \le \lambda_i \le 1, \text{ if } w_i = 0. \end{cases}$$

The following method found in Ben-Israel and Greville [2, pp. 14 - 18] may be used to find $A^{(1)}$.

Any matrix A can be premultiplied by a finite number of nonsingular matrices E_i representing elementary row operations and post multiplied by a finite number of nonsingular matrices P_i representing elementary column operations such that

$$EAP = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}$$

which implies that

$$A = E^{-1} \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1}$$

where A is of rank r, $E = E_m E_{m-1} \cdot \cdot \cdot E_1$, and $P = P_1 P_2 \cdot \cdot \cdot P_n$. E and P are nonsingular matrices since they were obtained by multiplying nonsingular matrices.

A generalized inverse of A satisfying AXA = A can be computed from

$$A^{(1)} = P \begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix} E$$

where L is any matrix of dimension (n-r) x (m-r) since

$$AA^{(1)}A = E^{-1} \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1}P \begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix} EE^{-1} \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1}$$

$$= E^{-1}\begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}\begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix}\begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1} \quad ;$$

$$= E^{-1}\begin{bmatrix} I_r & KL \\ 0 & 0 \end{bmatrix}\begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1} \quad ;$$

$$= E^{-1}\begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1} = A.$$

It should be noted that if $m = r$ and or $n = r$ special consideration should be given to $\begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix}$. In particular, L vanishes, so that the matrix becomes $\begin{bmatrix} I_r & 0 \end{bmatrix}$ if $m = r$ or $\begin{bmatrix} I_r \\ 0 \end{bmatrix}$ if $n = r$.

The columns of the $n \times (n-r)$ matrix $P\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$ form a basis of N(A) since

i. $AP\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix} = E^{-1}\begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} P^{-1}P\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix} = 0$

and hence is contained in N(A);

ii. By Theorem 8, the rank of $P\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$ equals the rank of $\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$ since P is nonsingular. The rank of $\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$ is n-r. Therefore the rank of $P\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$ is n-r. Thus the columns of $P\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$ are linearly independent. By Theorem 7 the dimension of N(A) is n-r. Thus the matrix $P\begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$

is a linearly independent subset of N(A) with n-r elements. Any such set is automatically a basis of N(A) [4, p. 69].

EXAMPLE:

Maximize $2x_1 - x_2 - x_3 + 3x_4$

subject to

$$0 \leq x_1 + 2x_2 - x_3 \leq 1$$
$$-3 \leq -x_1 + x_3 - x_4 \leq 0$$
$$1 \leq 2x_1 + x_2 - 3x_3 + x_4 \leq 3.$$

Solution: From the equations,

$$a = \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \quad c = \begin{bmatrix} 2 \\ -1 \\ -1 \\ 3 \end{bmatrix}, \quad \text{and } A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ -1 & 0 & 1 & -1 \\ 2 & 1 & -3 & 1 \end{bmatrix}.$$

To find $A^{(1)}$, A must first be written in the form

$$EAP = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}.$$

It is found that

$$E = \begin{bmatrix} 1/2 & -5/2 & -1 \\ 1/2 & 1/2 & 0 \\ 1/2 & -3/2 & -1 \end{bmatrix}, \quad P = I_4, \text{ and}$$

$$EAP = \begin{bmatrix} 1/2 & -5/2 & -1 \\ 1/2 & 1/2 & 0 \\ 1/2 & -3/2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 0 \\ -1 & 0 & 1 & -1 \\ 2 & 1 & -3 & 1 \end{bmatrix};$$

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

Using $P \begin{bmatrix} -K \\ I_{n-r} \end{bmatrix}$, a basis of N(A) is $\begin{bmatrix} -3/2 \\ 1/2 \\ -1/2 \\ 1 \end{bmatrix}$, and the IP is bounded since c is perpendicular to N(A).

Having defined

$$A^{(1)} = P \begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix} E ,$$

$$A^{(1)} = I_4 \begin{bmatrix} I_3 \\ 0 \end{bmatrix} \begin{bmatrix} 1/2 & -5/2 & -1 \\ 1/2 & 1/2 & 0 \\ 1/2 & -3/2 & -1 \end{bmatrix};$$

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

Using the definition of η (u,v,w) and realizing that $u = a$, $v = b$, and $w = A^{(1)T}c$, the optimal solution can be computed.

$$w = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 0 \\ -5/2 & 1/2 & -3/2 & 0 \\ -1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \\ -1 \end{bmatrix}$$

so that $w_1 = 0$, $w_2 = -4$, and $w_3 = -1$. Thus

$$\eta(a,b,A^{(1)T}c) = \begin{bmatrix} \lambda_1(0) + (1 - \lambda_1)1 \\ -3 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 - \lambda_1 \\ -3 \\ 1 \end{bmatrix} .$$

Let $\beta_1 = (q - \lambda_1)$. Since by definition $0 \leq \lambda_1 \leq 1$, then $0 \leq (1 - \lambda_1) \leq 1$. Thus $0 \leq \beta_1 \leq 1$.

Therefore the optimal solution is

$$x = \begin{bmatrix} 1/2 & -5/2 & -1 \\ 1/2 & 1/2 & 0 \\ 1/2 & -3/2 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ -3 \\ 1 \end{bmatrix} + y$$

which can be rewritten

$$x = \beta_1 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 0 \end{bmatrix} - 3 \begin{bmatrix} -5/2 \\ 1/2 \\ -3/2 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} -3/2 \\ 1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

where γ is arbitrary.

CHAPTER VII

MARKOV CHAINS

A stochastic process in probability theory is " a family of random variables describing an empirical process, the development of which in time is governed by probabilistic law" [11, p. 516]. Stochastic processes have applications in many fields. For example, Markov chains, a special class of stochastic processes, are applicable to problems in sociology, genetics, and engineering.

> A Markov chain may be defined as a random process, the development of which is treated as a series of transitions between certain values, called the states, of the process, which are finite or countably infinite and which possess the property that the future probabilistic behavior of the process depends only on the present (given) state and not on the method by which the process arrived in that state [11, p. 516].

All matrices, vectors, and scalars in this chapter are taken over the field of real numbers.

DEFINITION 9: The n x n matrix $P = (p_{ij})$ is said to be a stochastic matrix if and only if $P \geq 0$ and

$$\sum_{j=1}^{n} p_{ij} = 1, \quad i = 1, 2, \ldots, n.$$

DEFINITION 10: A stochastic matrix is ergodic if one eigenvalue equals unity and all the other eigenvalues have magnitudes less than unity.

DEFINITION 11: A stochastic matrix P is <u>regular</u> if there exists a positive integer n such that P^n has all positive terms.

Let the column vector p(0) contain the initial probabilities of the various states. Since p(0) is a probability vector the sum of the elements of the vector is 1. Let the stochastic matrix P consist of rows which are probability vectors giving the probabilities of transition from one state to another. Given p(0) and P it is of particular interest in many cases to find p(n), the probability vector giving the probability that, at a particular time n, the process is at a certain state. The probability vector of a system at time n is given by

$$p^T(n) = p^T(0)P^n.$$

If P is ergodic or regular, as n becomes infinitely large, the sequence $\{p(n)\}$ converges to a unique probability vector $p(\infty)$ called the <u>stationary vector</u> [11, p. 569] . The stationary vector, $p(\infty)$, satisfies

$$p^T(\infty)P = p^T(\infty)$$

and

$$p^T(\infty)e = 1,$$

where e is a vector having 1 in each component.

It is in the computation of $p(\infty)$ that the generalized inverse may be used. The following theorems are stated without proof and serve as a computational method for finding $p(\infty)$. For a reference to the two theorems see [1, p. 262].

THEOREM 21: If P is a regular or an ergodic stochastic matrix and

$$y^T = e^T\left[I - (P - I)(P - I)^+\right]$$

then $\left(\frac{1}{e^T y}\right) y^T$ is the unique stationary vector for P.

The preceeding theorem allows P to be either regular or ergodic but also limits the generalized inverse used to be the Moore-Penrose inverse. However, if P is required to be ergodic the following theorem allows the use of any generalized inverse satisfying at least one of the Moore-Penrose properties, equations (3), (4), (5), and (6).

THEOREM 22: Let P be an erogodic stochastic matrix and define

$$M = I - (P - I)(P - I)^-$$

where $(P - I)^-$ is any generalized inverse of $(P - I)$ satisfying at least one of the four Moore-Penrose properties, equations (3), (4), (5), and (6). If y^T is any nonzero row of M then

$$\left(\frac{1}{e^T y}\right) y^T = p^T(\infty)$$

is the unique stationary vector.

EXAMPLE: [11, p. 521]. Four quarterbacks are warming up by throwing a football to one another. Let i_1, i_2, i_3, and i_4 denote the four quarterbacks. It has been observed that i_1 is as likely to throw the ball to i_2 as to i_3 and i_4. Player i_2 never throws to i_3 but splits his throws between i_1 and i_4. Quarterback i_3 throws twice as many passes to i_1 as to i_4 and never throws to i_2, but i_4 throws only to i_1. This process forms a Markov chain because the player who is about to throw the ball is not influenced by the player who had the ball before him.

Use the Moore-Penrose inverse to find the stationary vector $p(\infty)$.

Solution: The transition probability matrix is

$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 0 & 1/2 \\ 2/3 & 0 & 0 & 1/3 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

For $n = 4$, P^n is positive, so by definition P is regular. Since P is regular it is possible to find $p(\infty)$ using Theorem 21. Solving for y^T,

$$y^T = e^T \left[I - (P - I)(P - I)^+\right];$$

$$= e^T \left[I - \begin{bmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/2 & -1 & 0 & 1/2 \\ 2/3 & 0 & -1 & 1/3 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/2 & -1 & 0 & 1/2 \\ 2/3 & 0 & -1 & 1/3 \\ 1 & 0 & 0 & -1 \end{bmatrix}^+ \right];$$

$$= e^T \left[I - \begin{bmatrix} -1 & 1/3 & 1/3 & 1/3 \\ 1/2 & -1 & 0 & 1/2 \\ 2/3 & 0 & -1 & 1/3 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} -.27998 & .15667 & .15668 & .28722 \\ .12041 & -.70987 & .29014 & .03192 \\ .05658 & .26886 & -.73114 & .15958 \\ .1028 & .28433 & .28433 & -.47871 \end{bmatrix} \right] ;$$

$$= e^T \left[I - \begin{bmatrix} .37331 & -.208896 & -.20890 & -.38295 \\ -.2089 & .93037 & -.06963 & -.12767 \\ -.2089 & -.06963 & .93037 & -.12767 \\ -.38298 & -.12766 & -.12765 & .76593 \end{bmatrix} \right] ;$$

$$= e^T \begin{bmatrix} .62669 & .208896 & .20890 & .38295 \\ .2089 & .069631 & .06963 & .12767 \\ .2089 & .069631 & .06963 & .12767 \\ .38298 & .12766 & .12765 & .23407 \end{bmatrix} ;$$

$$= \begin{bmatrix} 1.4275 & .47582 & .47582 & .87236 \end{bmatrix} .$$

Thus $\left(\frac{1}{e^T y} \right) y^T = \frac{1}{3.2515} \begin{bmatrix} 1.4275 & .47582 & .47582 & .87236 \end{bmatrix} ;$

$$= \begin{bmatrix} .43902 & .14634 & .14634 & .2683 \end{bmatrix} = p^T(\infty).$$

The probability vector $p(\infty)$ gives the long-range prediction regarding probabilities of the ball being thrown to each quarterback. Thus over a long period of time, quarterback 1 is likely to be thrown the ball about 44% of the time; quarterback 2, 15% of the time; quarterback 3, 15% of the time; and quarterback 4, about 27% of the time.

CHAPTER VIII

FIGURES

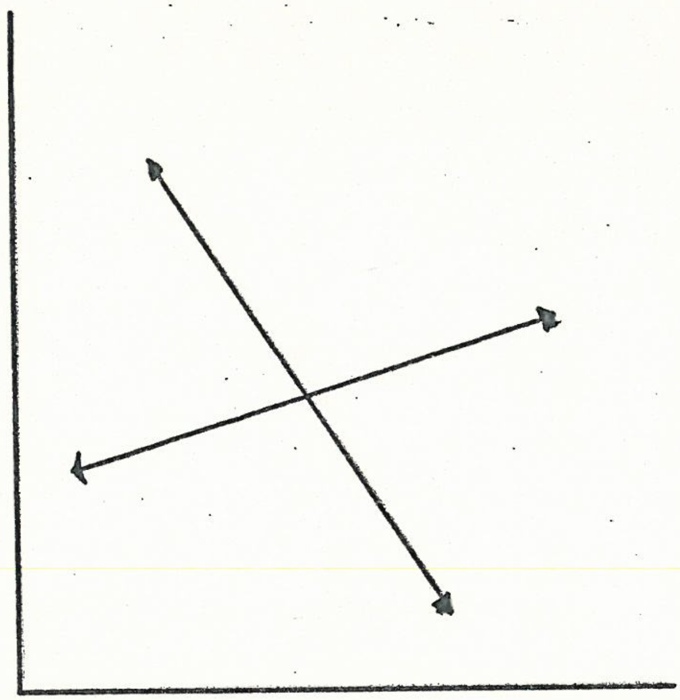

FIGURE 1 Intersecting lines

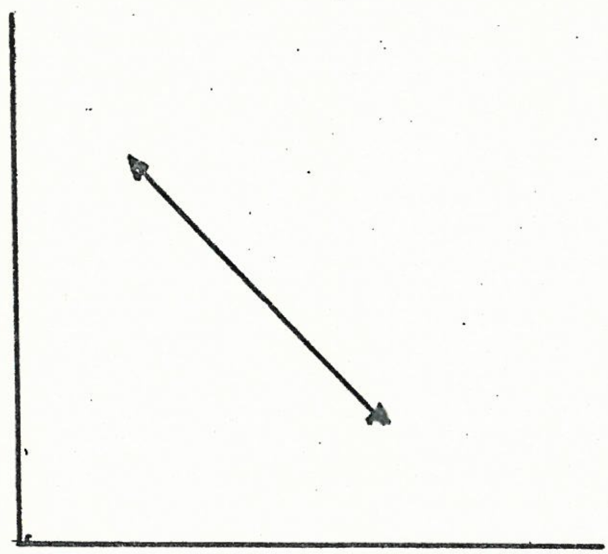

FIGURE 2 Coinciding lines

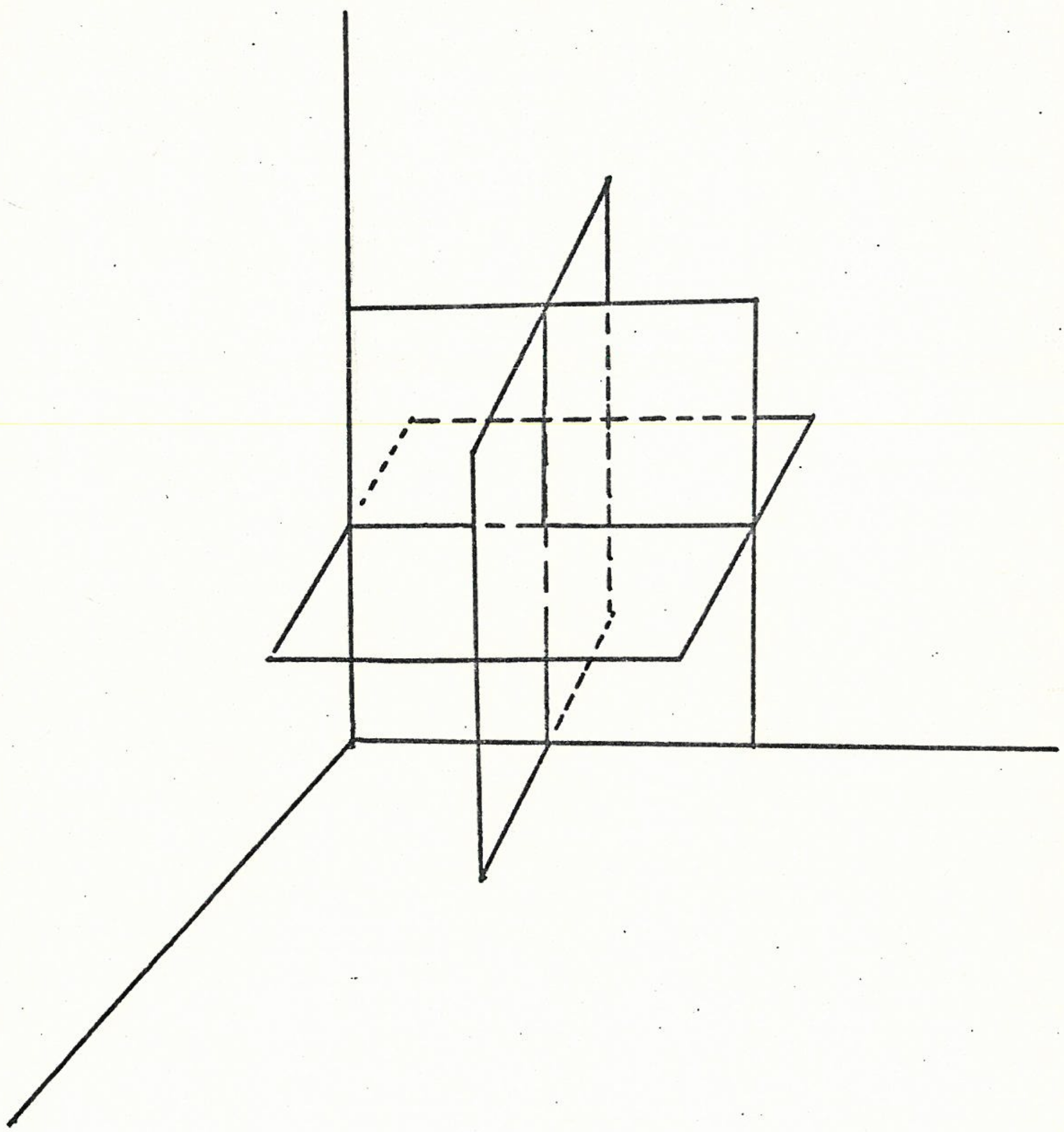

FIGURE 3 Three planes intersecting at a point

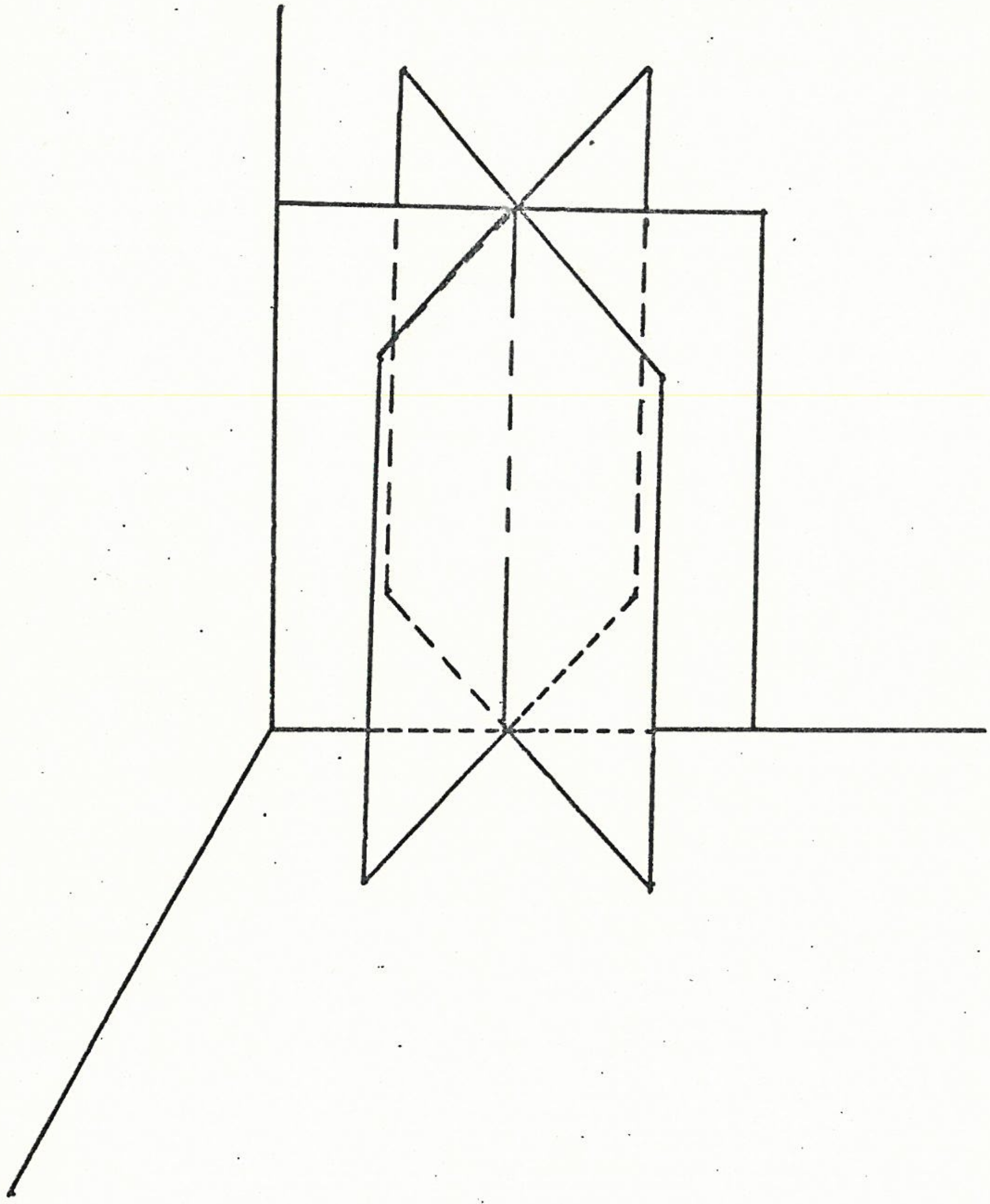

FIGURE 4 Three planes intersecting in a line

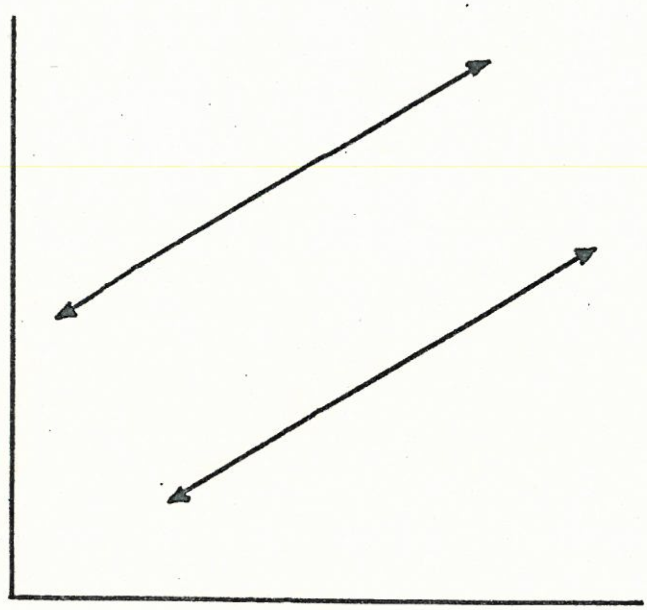

FIGURE 5 Parallel lines

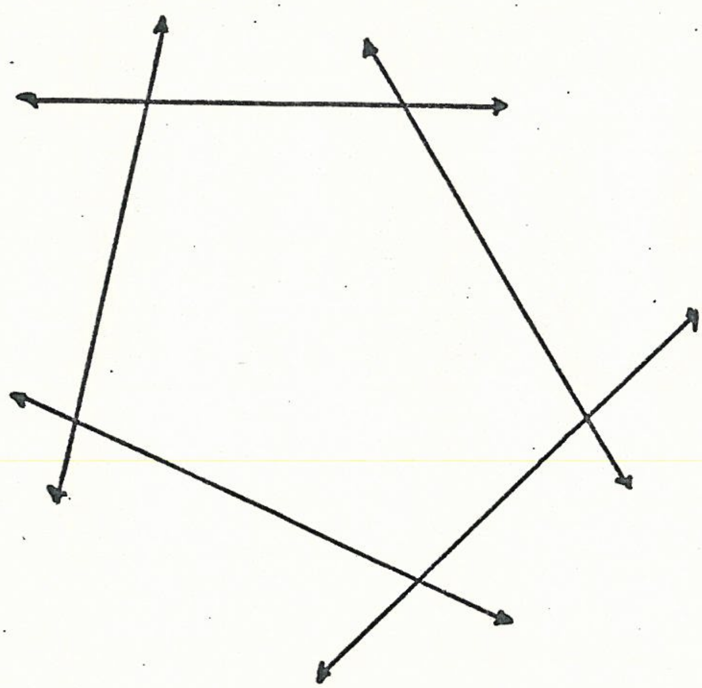

FIGURE 6a   Lines forming a polygon

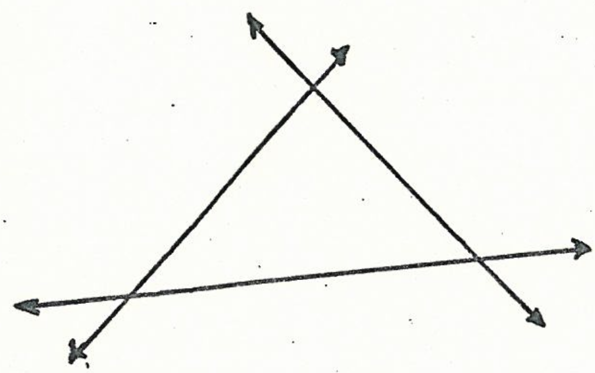

FIGURE 6b   Lines forming a triangle

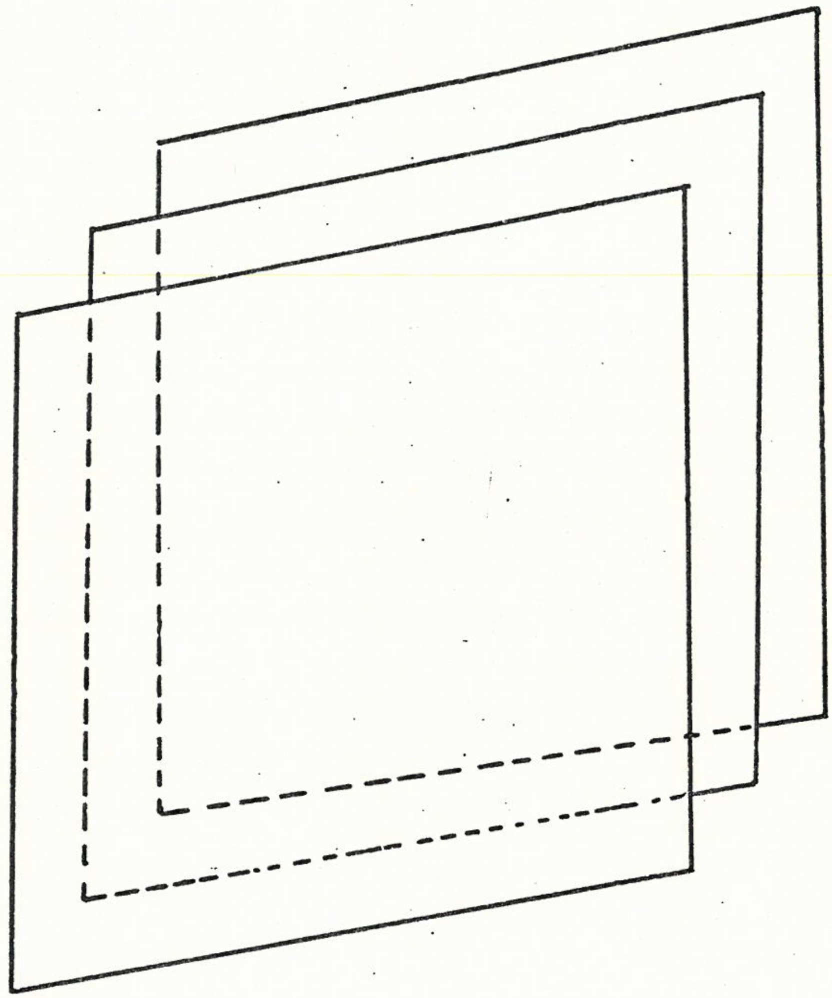

FIGURE 7 Three parallel planes

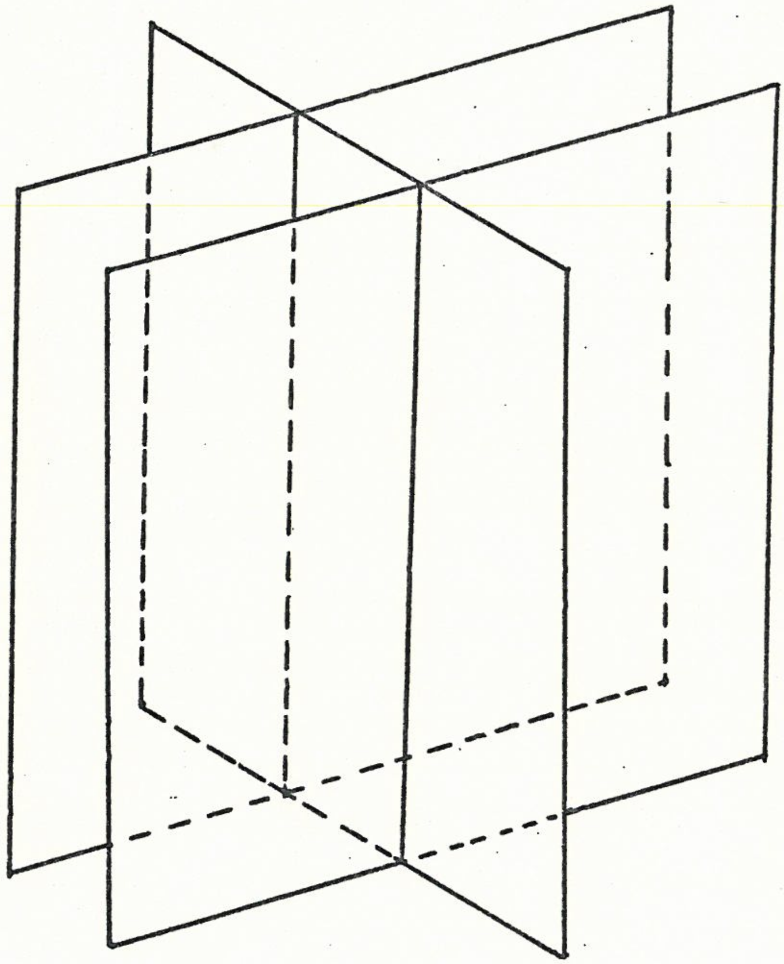

FIGURE 8 Two parallel planes intersected by a third plane

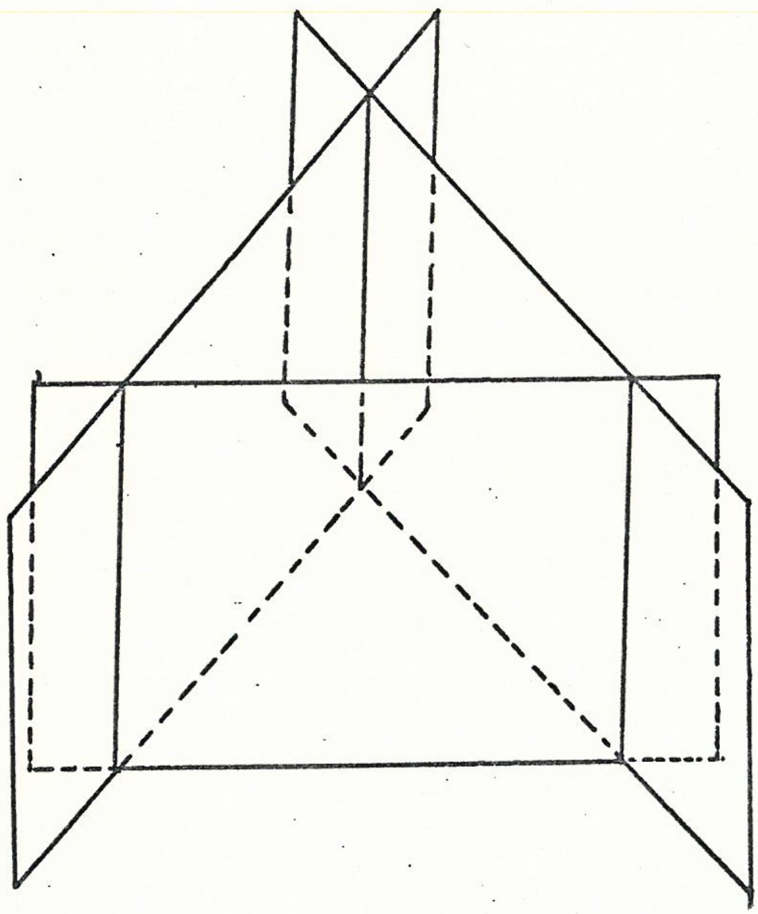

FIGURE 9 Three planes intersecting in three mutually parallel lines

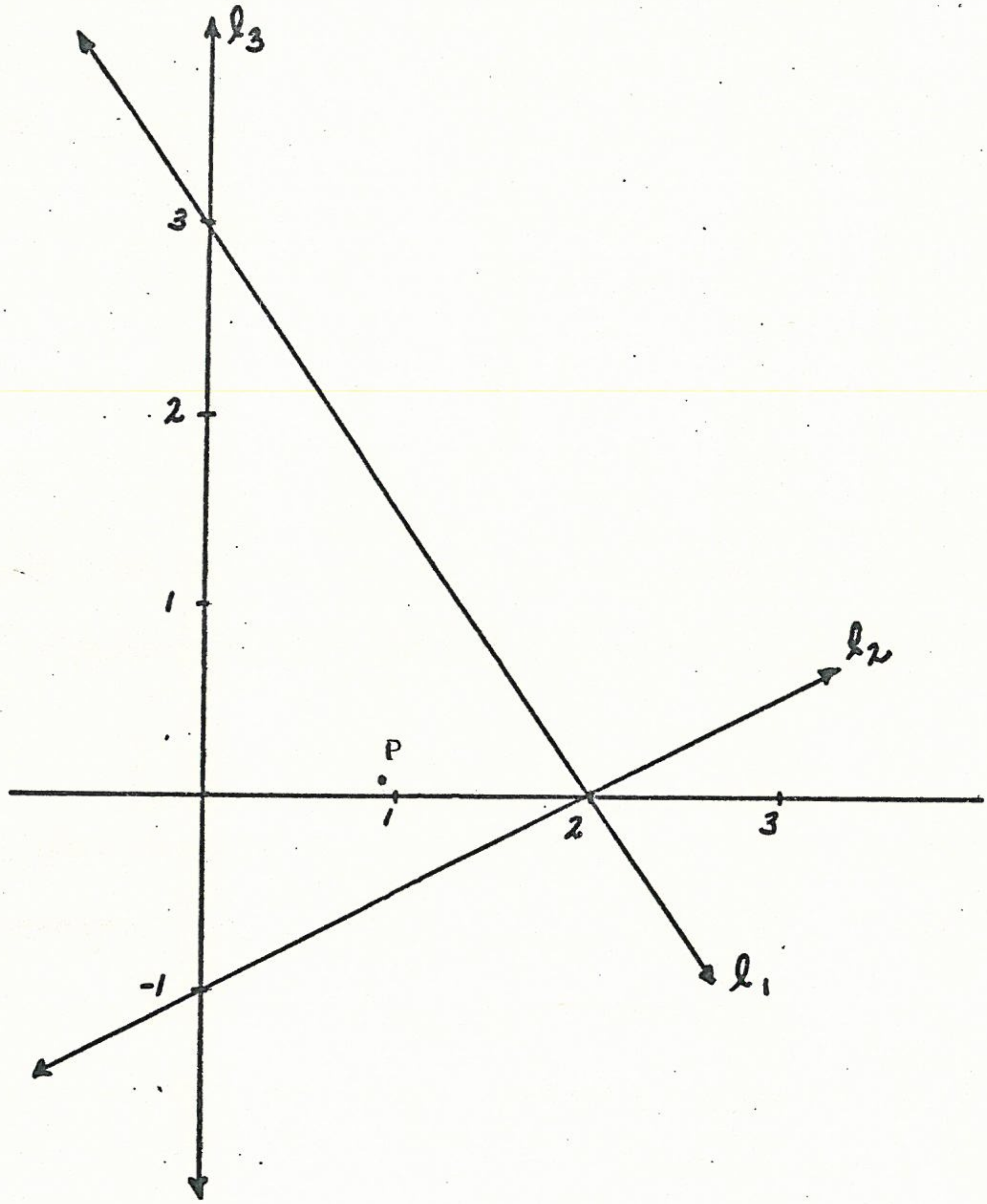

FIGURE 10 A graph of the equations.

l_1: $3x + 2y = 6$

l_2: $4x - 8y = 8$

l_3: $x = 0$.

Also plotted is the best approximate solution, $P = (.94, .06)$.

BIBLIOGRAPHY

1. Thomas L. Boullion and Patrick L. Odell, eds., Proceedings of the Symposium on Theory and Applications of Generalized Inverses of Matrices, Texas Technological College Mathematics Series No. 4. Texas Tech. Press, Lubbock, Texas, 1968.

2. Adi Ben-Israel and Thomas N. E. Greville, Generalized Inverses: Theory and Applications, John Wiley & Sons, Inc., New York, 1974.

3. F. E. Hohn, Elementary Matrix Algebra, Macmillan Publishing Company, New York, 1970.

4. Bernard Kolman, Elementary Linear Algebra, Macmillan Publishing Company, Inc., New York, 1970.

5. Peter Lancaster, Theory of Matrices, Academic Press, New York, 1969.

6. Seymour Lipschutz, Schaum's Outline of Theory and Problems of Linear Algebra, Schaum's Outline Series, McGraw-Hill Book Company, New York, 1968.

7. E. H. Moore, "On the Reciprocal of the General Algebraic Matrix (abstract)", Bull. Amer. Math. Soc. 26 (1920), 394-395.

8. R. Penrose, "A Generalized Inverse for Matrices", Proc. Cambridge Philos. Soc. 51 (1955), 406-413.

9. Sam Perlis, Theory of Matrices, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1952.

10. C. Radhakrishna Rao and Sujit Kumar Mitra, Generalized Inverse of Matrices and Its Applications, John Wiley & Sons, Inc., New York, 1971.

11. Chris P. Tsokos, Probability Distributions: An Introduction to Probability Theory with Applications, Duxbury Press, Belmont, California, 1972.

12. Gareth Williams, A Course in Linear Algebra, Gordon and Breach Science Publishers, New York, 1972.

VITA

Eleanor Louise White was born in Richmond, Virginia, on October 2, 1955. She attended elementary school at Hanover Academy in Ashland, Virginia. Her secondary education found her attending Patrick Henry High School in Hanover County, Virginia, Bowling Green Senior High School in Caroline County, Virginia, and finally Hermitage High School in Henrico County, Virginia. She graduated from Hermitage High School in June, 1973. The following August she entered Longwood College in Farmville, Virginia. She is a member of Lychnos Society and Pi Mu Epsilon.