GENERALIZED INVERSE OF A MATRIX AND ITS APPLICATIONS

C. RADHAKRISHNA RAO
and
SUJIT KUMAR MITRA
INDIAN STATISTICAL INSTITUTE

1. Introduction

The concept of an inverse of a singular matrix seems to have been first introduced by Moore [1], [2] in 1920. Extensions of these ideas to general operators have been made by Tseng [3], [4], [5], but no systematic study of the subject was made until 1955 when Penrose [6], [7], unaware of the earlier work, redefined the Moore inverse in a slightly different way. About the same time one of the authors, Rao [8], gave a method of computing what is called a pseudoinverse of a singular matrix, and applied it to solve normal equations with a singular matrix in the least squares theory and to express the variances of estimators. The pseudoinverse defined by Rao did not satisfy all the restrictions imposed by Moore and Penrose. It was therefore different from the Moore–Penrose inverse, but was useful in providing a general theory of least squares estimation without any restriction on the rank of the observational equations. In a later paper, Rao [9] showed that an inverse with a much weaker definition than that of Moore and Penrose is sufficient in dealing with problems of linear equations. Such an inverse was called a generalized inverse (g inverse) and its applications were considered by Rao in [10], [11], [12], [13], and [14].

Some of the principal contributors to the subject since 1955 are Greville [15], Bjerhammer [16], [17], [18], Ben-Israel and Charnes [19], Chipman [20], [21], Chipman and Rao [22], and Scroggs and Odell [23]. Bose [24] mentions the use of g inverse in his lecture notes, "Analysis of Variance" [24]. Bott and Duffin [25] defined what is called a constrained inverse of a square matrix, which is different from a g inverse and is useful in some applications. Chernoff [26] considered an inverse of a singular nonnegative definite (n.n.d.) matrix, which is also not a g inverse but is useful in discussing some estimation problems.

The g inverse satisfying the weaker definition given by Rao [9] is not unique and thus presents an interesting study in matrix algebra. In a publication in 1967 [27], Rao showed how a variety of g inverses could be constructed to suit different purposes and presented a classification of g inverses. The work was later pursued by Mitra [28], [29], who introduced some new classes of g inverses, and Mitra and Bhimasankaram [30], [31]. Further applications of g inverses were considered in a series of papers, Mitra and Rao [32], [33], [34], and Rao [35].

In the present paper we discuss a calculus of g inverses and show how it provides an elegant tool for the discussion of the Gauss-Markov problem of linear estimation, multivariate analysis when the variables have a singular covariance matrix, maximum likelihood estimation when the information matrix is singular, and so forth.

A systematic development of the calculus of generalized inverses and their applications are given in a forthcoming book by the authors, entitled *Generalized Inverse of Matrices and Its Applications* (Wiley, 1971).

2. Generalized inverse of a matrix

If A is an $m \times m$ nonsingular matrix, then there exists an inverse A^{-1} with the property $AA^{-1} = A^{-1}A = I$. If A is an $m \times n$ rectangular matrix with rank $n \leqq m$ then $(A^*A)^{-1}$ exists, and defining $A_L^{-1} = (A^*A)^{-1}A^*$ we find that $A_L^{-1}A = I$. In such a case A_L^{-1} is called a left inverse of A. Similarly a right inverse of A exists if its rank is $m \leqq n$ with the property $AA_R^{-1} = I$. When A^{-1}, A_L^{-1}, or A_R^{-1} exists we can express a solution of the equation $Ax = y$ in the form $x = A^{-1}y$ or $A_L^{-1}y$, or $A_R^{-1}y$. When such inverses do not exist, can we represent a solution of the consistent equation $Ax = y$ (where A may be rectangular or a square singular) in the form $x = Gy$? If such a G exists, we call it a generalized inverse of A, and represent it by A^-.

We provide three equivalent definitions of a g inverse.

DEFINITION 2.1. *An $n \times m$ matrix G is said to be a g inverse of an $m \times n$ matrix A if $x = Gy$ is a solution to the equation $Ax = y$ for any y such that the equation $Ax = y$ is consistent.*

DEFINITION 2.2. *G is a g inverse of A if $AGA = A$.*

DEFINITION 2.3. *G is a g inverse of A if AG is idempotent and $R(AG) = R(A)$ or GA is idempotent and $R(GA) = R(A)$, where $R(\cdot)$ denotes the rank of the matrix.*

A matrix G satisfying any one of these definitions is denoted by A^- and is called a g inverse. The following theorems establish the existence of A^- and its applications in solving equations, obtaining projections, and so forth. The proofs of some of these theorems are omitted as they are contained in Rao [27], and proofs of other theorems will appear in the forthcoming book by the authors, already cited.

THEOREM 2.1. *Let A be an $m \times n$ matrix. Then A^- exists. The entire class of g inverses is generated from any given inverse A^- by the formula*

$$A^- + U - A^-AUAA^- \tag{2.1}$$

where U is arbitrary, or by the formula

$$A^- + V(I - AA^-) + (I - A^-A)W \tag{2.2}$$

where V and W are arbitrary. Further a matrix is uniquely determined by the class (2.1) *or* (2.2) *of its g inverses.*

THEOREM 2.2. *Let $Ax = y$ be a consistent equation and A^- be a g inverse of A.*

(i) *Then $x = A^- y$ is a solution.*

(ii) *The class of all solutions is provided by $A^- y + (I - A^- A)z$, z arbitrary.*

(iii) *Let q be an n vector. Then $q'x$ has a unique value for all solutions x of $Ax = y$ if and only if $q' = q'A^- A$ or $q \in \mathscr{M}(A')$, the vector space generated by the columns of A'.*

THEOREM 2.3. *Let A be an $m \times n$ matrix and $\mathscr{M}(A) \subset \mathscr{E}^m$. The projection operator P onto $\mathscr{M}(A)$ can be expressed in the form*

$$P = A(A^*MA)^- A^*M, \tag{2.3}$$

where the inner product in $\mathscr{E}^m$ is defined as $(y, x) = x^*My$, M being a positive definite matrix and $(A^*MA)^-$ is any g inverse of A^*MA. Further P is unique for any choice of $(A^*MA)^-$.

It would be useful to recognize the situations in which a g inverse behaves like a regular inverse. Theorem 2.4 contains the main result in this direction.

THEOREM 2.4. *A necessary and sufficient condition that $BA^-A = B$ is that $B = DA$ for some D. Similarly for $B = AA^-B$ to hold, it is necessary and sufficient that $B = AD$.*

The following results are consequences of Theorem 2.4:

(a) $A(A^*A)^-(A^*A) = A$;

(b) $(A^*A)(A^*A)^-A^* = A^*$;

(c) $A(A^*VA)^-(A^*VA) = A$ and $(A^*VA)(A^*VA)^-A^* = A^*$ for any matrix V such that $R(A^*VA) = R(A)$;

(d) $A(A^*VA)^-A^*$ is invariant for any choice of $(A^*VA)^-$ and is of rank equal to the rank of A if $R(A^*VA) = R(A)$. Further, $A(A^*VA)^-A^*$ is hermitian if A^*VA is hermitian.

We provide a decomposition theorem involving g inverses of matrices which has a number of applications.

Let A be a matrix of order $m \times n$, and let A_i, B_i be matrices of order $m \times p_i$, $n \times q_i$, $i = 1, \cdots, k$. Write $A = (A_1 \vdots \cdots \vdots A_k)$ and $B = (B_1 \vdots \cdots \vdots B_k)$. Consider the following statements:

$$A_i^*\Lambda B_j = 0 \qquad \text{for all} \quad i \neq j. \tag{2.4}$$

$$G = \sum_i B_i(A_i^*\Lambda B_i)^- A_i^* \tag{2.5}$$

is a g inverse of Λ where $(A_i\Lambda B_i)^-$ is any g inverse.

THEOREM 2.5. (i) *Statement* (2.4) *implies statement* (2.5) *if and only if* $R(A^*\Lambda B) = R(\Lambda)$.

(ii) *Statement* (2.5) *implies statement* (2.4) *if and only if* $\Sigma R(A_i^*\Lambda) = \Sigma R(\Lambda B_i) = R(\Lambda)$.

An interesting corollary to Theorem 2.5 is the following.

COROLLARY 2.1. *Let A_i be an $m \times p_i$ matrix of rank r_i, $i = 1, \cdots, k$ such that $\Sigma r_i = m$. Further, let Λ be a positive definite (p.d.) matrix. Then the following two statements are equivalent*:

$$A_i^* \Lambda A_j = 0 \qquad \text{for all} \quad i \neq j, \tag{2.6}$$

$$\Lambda^{-1} = \Sigma A_i (A_i^* \Lambda A_i)^- A_i^*. \tag{2.7}$$

The true inverse of a nonsingular square matrix has the property that the inverse of the inverse is equal to the original matrix. This may not hold for any g inverse as defined in this section. We shall however show that a subclass of g inverses possesses an analogous property. We give the following definition.

DEFINITION 2.4. *An $n \times m$ matrix G is said to be a reflexive g inverse of an $m \times n$ matrix A if*

$$AGA = A \quad \text{and} \quad GAG = G. \tag{2.8}$$

We use the notation A_n^- to denote a reflexive g inverse.

THEOREM 2.6. *Any two of the following conditions imply the third*

(i) $A = AGA$,

(ii) $G = GAG$,

(iii) $R(G) = R(A)$.

For a proof of this theorem see Mitra [27]. It is seen that a reflexive g inverse could be equivalently defined by any two of the conditions (i), (ii) and (iii) in the theorem. Frame [36] uses the term *semi-inverse* to denote a matrix G obeying (i) and (iii).

3. Three basic types of g inverses

3.1. *Mathematical preliminaries and notations.* Let $\mathscr{E}^m$ represent an m dimensional vector space furnished with an inner product. The symbol (x, y) is used to denote the inner product between vectors x and y. The norm of a vector x is denoted by $\|x\| = [(x, x)]^{1/2}$.

Let A be an $m \times n$ matrix mapping vectors of $\mathscr{E}^n$ into $\mathscr{E}^m$. The adjoint of A denoted by $A^\#$ is defined by

$$(Ax, y)_m = (x, A^\# y)_n \tag{3.1}$$

where $(\cdot, \cdot)_m$ and $(\cdot, \cdot)_n$ denote inner products in $\mathscr{E}^m$ and $\mathscr{E}^n$, respectively. If $(y, x)_m = x^* M y$ and $(y, x)_n = x^* N y$ where M and N are positive definite matrices and $*$ denotes the conjugate transpose of a matrix, then relation (3.1) reduces to

$$x^* A^* M y = x^* N A^\# y \Rightarrow N A^\# = A^* M. \tag{3.2}$$

If A is an $m \times m$ square matrix mapping E^m into E^m, then $MA^\# = A^* M$.

We denote P_B the projection operator onto the space $\mathscr{M}(B)$ generated by the columns of B. It is characterized by the conditions:

(a) it is idempotent $P_B P_B = P_B$;

(b) it is selfadjoint $P_B = P_B^\#$.

If the inner product $(y, x)_m = x^* M y$, then condition (b) is equivalent to MP_B being hermitian.

3.2. *The g inverse for minimum norm solution.* It has been shown that $x = Gy$ is a solution of the consistent equation $Ax = y$ for any g inverse G of A (that is, satisfying the condition $AGA = A$), and the general solution is $x = Gy + (I - GA)z$ where z is arbitrary from, Theorem 2.2 (ii). We raise the question whether there exists a choice of G independently of y such that the solution Gy has a minimum norm in the class of all solutions of $Ax = y$. If such a G exists

$$\|Gy\| \leqq \|Gy + (I - GA)z\| \qquad \text{for all } z \text{ and } y \in \mathscr{M}(A), \tag{3.3}$$

that is,

$$\|GAx\| \leqq \|GAx + (I - GA)z\| \qquad \text{for all } z \text{ and } x. \tag{3.4}$$

This implies $\big(GAx, (I - GA)z\big) = 0$ for all z and x which implies in turn

$$(GA)^{\#}(I - GA) = 0 \quad \text{or} \quad (GA)^{\#} = (GA). \tag{3.5}$$

We now state the conditions for a g inverse G to provide a minimum norm solution of a consistent equation $Ax = y$.

THEOREM 3.1. *Let $Ax = y$ be a consistent equation and G be a g inverse of A such that Gy is a minimum norm solution. Then it is necessary and sufficient that any one of the following equivalent conditions is satisfied:*

(i) $AGA = A, (GA) = (GA)^{\#}$,

(ii) $AGA = A, (GA)^*N = N(GA)$ if $(y, x)_n = x^*Ny$,

(iii) $GA = P_{A^{\#}}$.

Condition (i) is already established and the equivalences of conditions (ii) and (iii) with (i) follow from the definitions of adjoint and projection operators.

We denote a g inverse which provides a minimum norm solution of $Ax = y$ by A_m^- or more explicitly $A_{m(N)}^-$, where N defines the inner product as in condition (ii), and refer to it as minimum N norm g inverse. Such an inverse exists; for example, $G = N^{-1}A^*(AN^{-1}A^*)^-$ satisfies the conditions of the theorem for any choice of the g inverse $(AN^{-1}A^*)^-$.

3.3. *The g inverse for a least square solution.* Let $Ax = y$ be an inconsistent equation in which case we seek a least squares solution by minimizing $\|Ax - y\|$. We raise the question whether there exists a matrix G such that $x = Gy$ is a least squares solution. If such a G exists

$$\|AGy - y\| \leqq \|Ax - y\| \quad \text{for all } x, y. \tag{3.6}$$

This implies $\big(Aw, (AG - I)y\big) = 0$ for all y, and $w = x - Gy$ implies $A^{\#}(AG - I) = 0$. Thus

$$AG = (AG)^{\#}, \qquad AGA = A. \tag{3.7}$$

THEOREM 3.2. *Let $Ax = y$ be a possibly inconsistent equation, and let G be a matrix such that Gy is a least squares solution of $Ax = y$. Then each of the following equivalent conditions is necessary and sufficient:*

(i) $AGA = A, \quad (AG) = (AG)^{\#}$,

(ii) $AGA = A, \quad (AG)^*M = M(AG)$ if $(y, x)_m = x^*My$,

(iii) $AG = P_A$.

Condition (i) is already established, and the equivalences of conditions (ii) and (iii) with (i) follow from the definitions. We denote a g inverse which provides a least squares solution of $Ax = y$ by A_l^- or more explicitly $A_{l(M)}^-$, where M defines the inner product as in condition (3.10), and refer to it as an M least squares g inverse. Such an inverse exists; for example, $G = (A^*MA)^-A^*M$ satisfies the condition of the theorem.

3.4. *The g inverse for minimum norm least squares solution.* A least squares solution of an inconsistent equation $Ax = y$ may not be unique, in which case we may seek for a matrix G such that Gy has minimum norm in the class of least squares solutions. If such a G exists,

$$\|Gy\|_n \leqq \|\xi\|_n, \{\xi: \|A\xi - y\|_m \leqq \|Ax - y\|_m \text{ for all } x\} \quad \text{for all } y \tag{3.8}$$

where $\|\cdot\|_m$ and $\|\cdot\|_n$ denote norms in $\mathscr{E}^m$ and $\mathscr{E}^n$, respectively. The condition (3.8) may be written

$$\|Gy\| \leqq \|\xi\|, \qquad \{\xi: A^\# A\xi = A^\# y\} \qquad \text{for all } y. \tag{3.9}$$

This implies $A^\# (I - AG) = 0$ and $G^\# (I - GA) = 0$ which in turn imply

$$AGA = A, (AG) = (AG)^\#, \qquad GAG = G, \qquad (GA) = (GA)^\#. \tag{3.10}$$

THEOREM 3.3. *Let $Ax = y$ be a possibly inconsistent equation and $x = Gy$ be a minimum norm least squares solution. Then each of the following equivalent conditions is necessary and sufficient:*

(i) $AGA = A$, $GAG = G$, $AG = (AG)^\#$, $GA = (GA)^\#$;

(ii) $AGA = A$, $GAG = G$, $(AG)^*M = MAG$, $(GA)^*N = NGA$, *when* $(y, x)_m = x^*My$ and $(y, x)_n = x^*Ny$;

(iii) $AG = P_A$, $GA = P_G$.

A matrix G satisfying any one of the above conditions is unique.

Conditions (i) are already established and the equivalences of (ii) and (iii) with (i) follow from the definitions. The uniqueness of G follows from the fact that a minimum norm solution of a linear equation is unique.

We denote the g inverse which provides a minimum norm least squares solution of $Ax = y$ by A^+ or more explicitly by A_{MN}^+, where M, N are matrices defining the inner products in $\mathscr{E}^m$, $\mathscr{E}^n$ as in condition (ii). Such an inverse exists; for example

$$G = A^\#A(A^\#AA^\#A)^-A^\# = A^\#(A^\#AA^\#)^-A^\# = P_{A^*}A^-P_A \tag{3.11}$$

satisfies the conditions of Theorem 3.3. We refer to A_{MN}^+ as the minimum N norm M least squares g inverse.

3.5. *Duality relationships between different g inverses.* An important theorem which establishes a duality relationship between minimum norm and least squares g inverses and which plays a key role in the Gauss-Markov theory of linear estimation is as follows.

THEOREM 3.4. *Let A be an $m \times n$ matrix and $(y, x)_m = x^*My$. Then*

$$(A^*)_{m(M)}^- = [A_{l(M^{-1})}^-]^*. \tag{3.12}$$

PROOF. Let G be a minimum M norm g inverse of A^*. Then using condition (ii) of Theorem 3.1,

$$(3.13) \qquad A^*GA^* = A^*, \qquad (GA^*)^*M = MGA^*.$$

Taking transposes and rewriting, (3.13) becomes

$$(3.14) \qquad AG^*A = A, \qquad (AG^*)^*M^{-1} = M^{-1}(AG^*)$$

which, using condition (ii) of Theorem 3.2, shows that G^* is an M^{-1} least squares g inverse of A. Then equation (3.12) is true. The duality result (3.12), in the special case when $M = I$, is also noted by Sibuya [37].

Another important theorem which has application in linear estimation is as follows.

THEOREM 3.5. *Let A be an $m \times n$ matrix and $(y, x)_m = x^*My$ and $(y, x)_n = x^*Ny$. Then,*

$$(3.15) \qquad (A^*)^+_{NM} = (A^+_{M^{-1}N^{-1}})^*.$$

The result follows from condition (ii) of Theorem 3.3.

The different types of g inverses considered in Sections 2 and 3 and the properties characterising them are given in Table I.

TABLE I

SOME TYPES OF g INVERSES

P_X denotes projection operator onto $\mathscr{M}(X)$, $\#$ denotes adjoint.

Symbol	Equivalent Conditions	Purpose
A^-	$AGA = A$	solving consistent equations
A_r^-	$AGA = A,\ GAG = G$	solving consistent equations
A_m^-	(i) $AGA = A,\ (GA)^\# = GA$ (ii) $GA = P_{A^\#}$	minimum norm solution
A_l^-	(i) $AGA = A,\ (AG)^\# = AG$ (ii) $AG = P_A$	least squares solution
A^+	(i) $AGA = A,\ GAG = G$ $(GA)^\# = GA,\ (AG)^\# = AG$ (ii) $AG = P_A,\ GA = P_G$	minimum norm least squares solution

In Theorems 3.1 to 3.3, we used norms defined by p.d. matrices M and N. We can extend the results to cases where M and N are n.n.d. matrices. In such a case we will be minimizing seminorms. Some results in this direction will appear in a forthcoming paper in Sankhyá.

3.6. *Singular value decomposition.* Let A be an $m \times n$ matrix of rank r and M and N are p.d. matrices of order m and n, respectively. Then A can be expressed in the form

$$(3.16) \qquad MAN = a_1\xi_1\eta_1^* + \cdots + a_r\xi_r\eta_r^*,$$

where $a_1^2, \cdots, a_r^2$ are the nonzero eigenvalues of A^*MA with respect to N^{-1}

or of ANA^* with respect to M^{-1}; ξ_i is the eigenvector of ANA^* with respect to M^{-1} corresponding to the eigenvalue a_i^2; and η_i is the eigenvector of A^*MA with respect to N^{-1} corresponding to the eigenvalue a_i^2. The representation (3.16) is called the singular value decomposition of A with respect to M and N. Using such a decomposition we can compute A_{MN}^+ as

$$A_{MN}^+ = a_1^{-1}\eta_1\xi_1^* + \cdots + a_r^{-1}\eta_r\xi_r^*. \tag{3.17}$$

4. Constrained inverse

Bott and Duffin [25] introduced what is called a constrained inverse of a square matrix, which is different from a g inverse, and considered its application in mechanics and in network theory. In this section we extend the concept of a constrained inverse to a general matrix and give some applications.

Let A be a matrix of order $m \times n$. $\mathscr{V}$ and $\mathscr{U}$ be subspaces in $\mathscr{E}^n$ and $\mathscr{E}^m$, respectively. In what follows we shall impose constraints of two different types to define a constrained inverse G of A.

CONSTRAINTS OF TYPE 1.

c: G maps vectors of $\mathscr{E}^m$ into $\mathscr{V}$;

r: *G^* maps vectors of $\mathscr{E}^n$ into $\mathscr{U}$.*

CONSTRAINTS OF TYPE 2.

C: *GA is an identity in $\mathscr{V}$,*

R: *$(AG)^*$ is an identity in $\mathscr{U}$.*

Inverses obtained by choosing various combinations of these constraints are listed below in Table II along with necessary and sufficient conditions for existence, and explicit forms, where F and E are matrices such that $\mathscr{V} = \mathscr{M}(E)$ and $\mathscr{U} = \mathscr{M}(F^*)$.

TABLE II

CONSTRAINED INVERSES OF VARIOUS TYPES

V and U are arbitrary matrices.

Notation	N.S. Condition for Existence	Algebraic Expression	Reference to Theorem
A_{cC}	$R(AE) = R(E)$	$E(AE)^-$	4.1
A_{rR}	$R(FA) = R(F)$	$(FA)^-F$	4.3
A_{cR}	$R(FAE) = R(F)$	$E(FAE)^-F + E[I - (FAE)^-FAE]U$	4.5
A_{rC}	$R(FAE) = R(E)$	$E(FAE)^-F + V[I - FAE(FAE)^-]F$	4.6
A_{crCR}	$R(FAE) = R(F) = R(E)$	$E(FAE)^-F$	4.7

THEOREM 4.1. *A_{cC} exists if and only if $R(AE) = R(E)$. In such a case A_{cC} is of the form $E(AE)^-$.*

PROOF. Using constraint c, $G = EX$ for some matrix X. Then constraint C gives

$$EXAE = E. \tag{4.1}$$

Equation (4.1) is solvable only if $R(AE) = R(E)$ in which case, (4.1) is equivalent to $AEXAE = AE$, or

$$X = (AE)^- \Rightarrow G = E(AE)^-. \tag{4.2}$$

The "if" part is trivial.

THEOREM 4.2. *A is a g inverse of A_{cC} but not necessarily the other way. A_{cC} is a g inverse of A if and only if $R(AE) = R(A)$.*

PROOF. Theorem 4.2 follows from Theorems 4.1 and 2.4. Theorems 4.3 and 4.4 follow on similar lines.

THEOREM 4.3. *A_{rR} exists if and only if $R(FA) = R(F)$. In such a case A_{rR} is of the form $(FA)^-F$.*

THEOREM 4.4. *A is a g inverse of A_{rR} but not necessarily the other way. A_{rR} is a g inverse of A if and only if $R(FA) = R(A)$.*

THEOREM 4.5. *A_{cR} exists if and only if $R(FAE) = R(F)$. In such a case A_{cR} is of the form*

$$E(FAE)^-F + E[I - (FAE)^-FAE]U, \tag{4.3}$$

where U is arbitrary.

PROOF. Using constraint c, $G = EX$, for some matrix X. Then constraint R gives

$$FAEX = F. \tag{4.4}$$

Equation (4.4) is solvable if and only if $R(FAE) = R(F)$, in which case a general solution is given by

$$X = (FAE)^-F + [I - (FAE)^-FAE]U, \tag{4.5}$$

where U is arbitrary. The "if" part is easy. Thus Theorem 4.5 is established. Theorem 4.6 can be proved on similar lines.

THEOREM 4.6. *A_{rC} exists if and only if $R(FAE) = R(E)$. In such a case A_{rC} is of the form,*

$$E(FAE)^-F + V[I - FAE(FAE)^-]F, \tag{4.6}$$

where V is arbitrary.

THEOREM 4.7. *A_{crCR} exists if and only if $R(FAE) = R(F) = R(E)$. In such a case A_{crCR} is unique and is given by the expression $E(FAE)^-F$.*

PROOF. The "if" part is trivial. The necessity of the rank condition follows as in Theorems 4.5 and 4.6. The uniqueness follows, since under the condition $R(FAE) = R(F) = R(E)$ both A_{cR} and A_{rC} are uniquely determined by the expression $E(FAE)^-F$. Look for example at the expression (4.3), for A_{cR} and check that when $R(FAE) = R(E)$,

$$FAE[I - (FAE)^-FAE] = 0 \Rightarrow E[I - (FAE)^-FAE] = 0. \tag{4.7}$$

NOTE 1. Let E_1 and F_1 be matrices such that $\mathcal{M}(E_1) = \mathcal{M}(E)$ and $\mathcal{M}(F_1) = \mathcal{M}(F)$, where F and E are as defined in Theorem 4.1. Then

$$R(AE) = R(E) \Rightarrow R(AE_1) = R(E_1), \tag{4.8}$$

$$R(FAE) = R(E) \Rightarrow R(F_1AE_1) = R(E_1), \tag{4.9}$$

$$E_1(F_1AE_1)^-F_1 = E(FAE)^-F, \tag{4.10}$$

so that A_{crCR} is unique for any choice of the matrices generating the subspaces $\mathscr{V}$ and $\mathscr{U}$.

NOTE 2. In particular let P and Q be projection operators onto $\mathscr{V}$ and $\mathscr{U}$, respectively. Then

$$A_{crCR} = P(QAP)^-Q. \tag{4.11}$$

NOTE 3. A is g inverse of A_{crCR} but the converse is true only under the additional condition $R(FAE) = R(A)$.

NOTE 4. When $\mathscr{V} = \mathscr{M}(A^\#)$ and $\mathscr{U} = \mathscr{M}(A)$, A_{crCR} coincides with A^+_{MN}. It may be of some historical interest to observe that Moore [1], [2] introduced his general reciprocal of a matrix as a constrained inverse of the type we are considering in this section.

Now we consider the special case where A is an $m \times m$ (square) matrix and the subspaces $\mathscr{V}$ and $\mathscr{U}$ are the same and discuss it in some detail. The constrained inverse G in such a case may be defined by the following conditions:

(a) G^* maps vectors of $\mathscr{E}^m$ into the subspace $\mathscr{V} \subset \mathscr{E}^m$;

(b) GA is an identity in $\mathscr{V}$.

This is a special case of A_{rC}, but we shall represent a matrix G satisfying the above two conditions by T, following the notation used by Bott and Duffin. (In condition (i) above Bott and Duffin used G instead of G^*, which does not characterize the matrix T used by them. Their definition leads to an inverse of the type A_{cC} which is not unique and so on.)

THEOREM 4.8. *Let E be a matrix such that $\mathscr{V} = \mathscr{M}(E)$. Then T exists if and only if $R(E^*AE) = R(E)$ in which case it is unique, and is of the form*

$$T = E(E^*AE)^-E^*. \tag{4.12}$$

Further T is independent of the choice of E.

The proof is on the same lines as in Theorem 4.7.

THEOREM 4.9. *Let P be the projection operator onto $\mathscr{V}$ and $R(PAP) = R(P)$. Then*

$$T = P(PAP)^-P = P(AP + I - P)^{-1}. \tag{4.13}$$

PROOF. The first part of equation (4.13) follows from Theorem 4.8 as we can choose E to be P. For the second part, it is easy to see that $(AP + I - P)$ is nonsingular and admits a regular inverse when $R(PAP) = R(P)$. Further

$$[P(PAP)^-P - P(AP + I - P)^{-1}](AP + I - P) = 0 \tag{4.14}$$

giving $P(PAP)^-P = P(AP + I - P)^{-1}$ which is the expression used by Bott and Duffin.

THEOREM 4.10. *Let A be an $m \times m$ matrix and T be the constrained inverse as obtained in Theorem* 4.8. *Then:*

(i) *any arbitrary vector h admits a unique decomposition $h = Au + w$, $u \in \mathscr{V}$ and $w \in \mathscr{V}^{\perp}$;*

(ii) *the quadratic function $Q = (v - e)^*A(v - e) - 2f^*v$, where e and f are given vectors, attains a stationary value for variations of v in $\mathscr{V}$. If A is an n.n.d. matrix, Q attains the minimum.*

PROOF. (i) Let $h = Au + w$. Multiplying by T on both sides $Th = TAu + Tw = u$. Then $w = h - ATh$. It is easily checked that $Th \in \mathscr{V}$ and $h - ATh \in \mathscr{V}^{\perp}$. Further, if $Au_1 + w_1$ is another decomposition, $0 = A(u - u_1) + w - w_1$. Multiplying both sides by T, $u - u_1 = 0$ and hence $w - w_1 = 0$, so that the decomposition is unique.

(ii) Substituting $v = v_0 + \delta$, $v_0 \in \mathscr{V}$, $\delta \in \mathscr{V}$, and retaining only linear terms in δ, the quadratic form becomes

$$(v_0 - e)^*A(v_0 - e) - 2f^*v_0 - 2\delta^*(f + Ae - Av_0). \tag{4.15}$$

Then v_0 is a stationary point if $\delta^* (Av_0 - f - Ae) = 0$ or $Av_0 + w = Ae + f = h$, say, where $w \in \mathscr{V}^{\perp}$. Applying result (i) of the theorem, v_0 exists and has the value $v_0 = Th = T(Ae + f)$.

To show that Q attains a minimum at v_0 when A is n.n.d., let us observe that for any $v \in \mathscr{V}$,

$$\begin{aligned}(v - e)^*A(v - e) - 2f^*v = (v_0 - e)^*A(v_0 - e) - 2f^*v_0 \\ + (v - v_0)^*A(v - v_0).\end{aligned} \tag{4.16}$$

This completes the proof of Theorem 4.10.

5. Method of least squares

We show how Theorem 3.4 expressing the duality between minimum norm and least squares inverses provides a simple and an elegant demonstration of the minimum variance property of least squares estimators in the Gauss-Markov model. It also shows how the least squares method comes in a natural way while seeking for minimum variance estimators.

The Gauss-Markov model is characterized by the triplet $(Y, X\beta, \Lambda)$ where Y is $n \times 1$ vector of random variables such that $E(Y) = X\beta$, $D(Y) = \Lambda$ (variance-covariance matrix of Y).

5.1. *Unbiased estimation.* Let $p'\beta$ be a parametric function where $p \in \mathscr{M}(X')$. We wish to find a linear function $L'Y$ of Y such that $E(L'Y) \equiv p'\beta$ and the variance $V(L'Y) = L'\Lambda L$ is a minimum. The condition on expectation gives that $L'X\beta \equiv p'\beta \Leftrightarrow X'L = p$. The equation $X'L = p$ is consistent and what we need is a minimum norm solution, norm being defined as $\|L\|^2 = L'\Lambda L$. The optimum value of L is obviously, using a minimum Λ norm g inverse

$$L = (X')^-_{m(\Lambda)}p, \tag{5.1}$$

giving the minimum variance linear estimator

$$L'Y = p'[(X')^{-}_{mm(\Lambda)}]'Y = p'X^{-}_{l(\Lambda^{-1})}Y = p'\hat{\beta} \tag{5.2}$$

using the duality Theorem 3.4, where $\hat{\beta}$ is the Λ^{-1} least squares solutions of the equation $Y = X\beta$, that is, which minimizes

$$\|Y - X\beta\|^2 = (Y - X\beta)^*\Lambda^{-1}(Y - X\beta). \tag{5.3}$$

5.2. *Minimum bias estimation.* If $p \notin \mathcal{M}(X')$, the parametric function $p'\beta$ does not admit an unbiased linear estimator. The magnitude of bias in $L'Y$ is $(X'L - p)'\beta$. The bias may be minimized by choosing L such that

$$\|X'L - p\|^2 = (X'L - p)'N(X'L - p) \tag{5.4}$$

is a minimum, where N is a specified positive definite matrix. Subject to minimum bias we wish to minimize the variance $L'\Lambda L$. The problem then is that of finding minimum Λ norm N least squares solution of the equation $X'L = p$. Then the optimum value of L is

$$L = (X')^{+}_{N\Lambda}p, \tag{5.5}$$

giving the least bias minimum variance linear estimator

$$L'Y = p'[(X')^{+}_{N\Lambda}]'Y = p'X^{+}_{\Lambda^{-1}N^{-1}}Y = p'\hat{\beta} \tag{5.6}$$

using Theorem 3.5, where $\hat{\beta}$ is the N^{-1} norm Λ^{-1} least squares solution of the equation $Y = X\beta$.

6. Maximum likelihood estimation when the information matrix is singular

Let $p(x, \theta)$ be the probability density, where x stands for observed data and θ for n unknown parameters $\theta_1, \cdots, \theta_n$. Then

$$L(\theta, x) = \log p(x, \theta) \tag{6.1}$$

as a function of θ for given x is known as the log likelihood of parameters. Let $f_i(x, \theta)$ or simply f_i be defined by

$$f_i = \frac{\partial L}{\partial \theta_i}, \qquad i = 1, \cdots, n, \tag{6.2}$$

and let the vector $(f_1, \cdots, f_n)'$ be f. The information matrix on θ is defined by

$$H = E(ff') \tag{6.3}$$

The maximum likelihood m.l. estimate of θ is usually obtained from the equation $f = 0$, and the asymptotic theory of estimation is well known when the matrix H is not singular in the neighborhood of the true value.

If $L(\theta, x)$ depends essentially on $s < n$ independent functions $\phi_1, \cdots, \phi_s$ of θ, then H becomes singular and not all the parameters $\theta_1, \cdots, \theta_n$ are estimable.

Only $\phi_1, \cdots, \phi_s$ and their functions are estimable. In such a case we can define the log likelihood $L(\theta, x)$ in terms of fewer parameters as $L(\phi, x)$ where $\phi' = (\phi_1, \cdots, \phi_s)$ such that J, the information matrix on ϕ, is nonsingular. Then the usual theory would apply. Of course, there is some arbitrariness in the choice of ϕ but this does not cause any trouble. However, the calculus of g inverses enables us to deal with the likelihood as a function of the original parameters and obtain their m.l. estimates and the associated asymptotic variance-covariance matrix. When all the parameters are not estimable, the individual estimates and the variance-covariance matrix so obtained are not meaningful, but they are useful in computing m.l. estimates and standard errors of estimable parametric functions. (We have learned from H. Rubin at the Symposium that he considered such an approach and obtained results similar to ours.)

6.1. *Method of scoring with a singular information matrix.* The m.l. estimates are obtained by solving the equations

$$f_i(x, \theta) = 0, \qquad i = 1, \cdots, n. \tag{6.4}$$

The equations (6.4) are usually complicated in which case one obtains solutions by successive approximations using a technique such as Fishers' method of scoring (see Rao [10], pp. 302–309). Let θ_0 be an approximate solution and $\delta\theta$ the correction. Then neglecting higher order terms in $\delta\theta$

$$-f(x, \theta_o) = H\delta\theta, \tag{6.5}$$

where H is computed at θ_0. Since H is singular, there is no unique solution to (6.5) and therefore, the question of choosing a suitable solution arises. A natural choice is a solution with a minimum norm

$$\delta\theta = -H_m^- f(x, \theta_0) \tag{6.6}$$

We may terminate the iterative procedure when the correction needed is negligible. Let $\hat{\theta}$ be the approximate solution thus obtained and H^- any g inverse of H computed at $\hat{\theta}$. As observed earlier $\hat{\theta}$ and H^- are not meaningful when H is singular.

A parametric function $\psi(\theta)$ is said to be estimable if ξ_θ, the vector of derivatives of $\psi(\theta)$ with respect to $\theta_1, \cdots, \theta_n$, belongs to $\mathcal{M}(H)$. For such a function $\psi(\theta)$, $\psi(\hat{\theta})$ is the unique m.l. estimate for any choice $\hat{\theta}$ of m.l. estimate of θ and the asymptotic variance of $\psi(\hat{\theta})$ is

$$\xi_\theta' H^- \xi_\theta \tag{6.7}$$

which is unique for any choice of the g inverse of H.

Chernoff [26] defined in inverse of a singular information matrix, which is not a g inverse in our sense. For instance, when no individual parameter is estimable, Chernoff's inverse does not exist (all the entries become infinite), while H^- exists and can be used as in formula (6.7) to find standard errors of estimable parametric functions.

7. Distribution of quadratic functions in normal variates

In this section we shall study the distribution of a quadratic function $Y'AY + 2b'Y + c$ in normally distributed variables $Y_1, Y_2, \cdots, Y_n$ and obtain conditions under which such a function would have a chi square (χ^2) distribution (central or noncentral). We denote a central χ^2 distribution with k degrees, of freedom by $\chi^2(k)$ and the noncentral distribution with parameter δ by $\chi^2(k, \delta)$. Also we denote a p variate normal distribution by $N_p(\mu, \Sigma)$ where μ is the mean vector and Σ is the dispersion matrix which may be singular (see Rao [10], p. 437).

THEOREM 7.1. *Let* $Y \sim N_n(\mu, I)$. *Then*

$$\Sigma\lambda_i Y_i^2 + 2\Sigma b_i Y_i + c \sim \chi^2(k, \delta) \tag{7.1}$$

if and only if

(i) *each* λ_i *is either* 0 *or* 1,

(ii) $b_i = 0$ *if* $\lambda_i = 0$, *and*

(iii) $c = \Sigma b_i^2$,

in which case the number of degrees of freedom is $k = \Sigma\lambda_i$ *and the noncentrality parameter* $\delta = \Sigma\lambda_i(\mu_i + b_i)^2$.

PROOF. The theorem is easy to establish by comparing the characteristic function of $\Sigma_{i=1}^n \lambda_i Y_i^2 + 2\,\Sigma_{i=1}^n b_i Y_i + c$ and of $\Sigma_{i=1}^k (X_i + \nu_i)^2$ where the X_i are independent standard univariate normal variables.

THEOREM 7.2. *Let* $Y \sim N_n(\mu, I)$. *Then*

$$Y'AY + 2b'Y + c \sim \chi^2(k, \delta) \tag{7.2}$$

if and only if

(i) $A^2 = A$,

(ii) $b \in \mathscr{M}(A)$, *and*

(iii) $c = b'b$, *in which case the d.f.*, $k = R(A) = \operatorname{tr} A$ *and*

$$\delta = (b + \mu)'A(b + \mu). \tag{7.3}$$

PROOF. There exists an orthogonal matrix P such that $A = P'\Delta P$ where Δ is diagonal. Under the transformation $Z = PY$

$$Y'AY + 2b'Y + c = Z'\Delta Z + 2(Pb)'Z + c. \tag{7.4}$$

Further $Z \sim N_n(P\mu, I)$. Hence by Theorem 7.1

$$Y'AY + 2b'Y + c \sim \chi^2(k, \delta) \tag{7.5}$$

if and only if

(i) each diagonal element of Δ is either 0 or 1, that is, $\Delta^2 = \Delta$ or equivalently $A^2 = A$;

(ii) the ith coordinate of Pb is 0 if the ith diagonal element of Δ is 0, that is, $Pb \in \mathscr{M}(\Delta)$ is equivalent to $b \in \mathscr{M}(A)$; and

(iii) $c = (Pb)'Pb = b'b$.

Check that $k = \operatorname{tr} \Delta = \operatorname{tr} A = R(A)$ and

$$\delta = (Pb + P\mu)'\Delta(Pb + P\mu) = (b + \mu)'A(b + \mu). \tag{7.6}$$

THEOREM 7.3. *Let $Y \sim N_n(\mu, \Sigma)$ where Σ could be singular. Then*

$$Y'AY + 2b'Y + c \sim \chi^2(k, \delta) \tag{7.7}$$

if and only if

(i) $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$ *or equivalently* $(\Sigma A)^3 = (\Sigma A)^2$,

(ii) $\Sigma(A\mu + b) \in \mathscr{M}(\Sigma A \Sigma)$,

(iii) $(A\mu + b)'\Sigma(A\mu + b) = \mu'A\mu + 2b'\mu + c$, *in which case* $k = \operatorname{tr} A\Sigma$, $\delta = (b + A\mu)'\Sigma A \Sigma(b + A\mu)$.

PROOF. We express $Y = \mu + FZ$, where F is an $n \times r$ matrix of rank r such that $\Sigma = FF'$ and $Z \sim N_r(0, I)$. In terms of Z we have

$$Y'AY + 2b'Y + c = Z'F'AFZ + 2(A\mu + b)'FZ + \mu'A\mu + 2b'\mu + c. \tag{7.8}$$

Applying Theorem 7.2 to the quadratic function in Z we have therefore the following necessary and sufficient conditions for $\chi^2(k, \delta)$ distribution:

(i) $(F'AF)^2 = F'AF \Leftrightarrow \Sigma A \Sigma A \Sigma = \Sigma A \Sigma \Leftrightarrow (\Sigma A)^3 = (\Sigma A)^2$,

(ii) $F'(A\mu + b) \in \mathscr{M}(F'AF) \Leftrightarrow \Sigma(A\mu + b) \in \mathscr{M}(\Sigma A \Sigma)$, and

(iii) $(A\mu + b)'\Sigma(A\mu + b) = \mu'A\mu + 2b'\mu + c$.

Observe that $k = \operatorname{tr} F'AF = \operatorname{tr} A\Sigma$ and

$$\delta = (A\mu + b)'FF'AFF'(A\mu + b) = (A\mu + b)'\Sigma A \Sigma(A\mu + b) \tag{7.9}$$

COROLLARY 7.1. *Let $Y \sim N_n(\mu, \Sigma)$. Then $Y'\Sigma^- Y \sim \chi^2(k, \delta)$ if and only if $\mu'(\Sigma^-\Sigma\Sigma^- - \Sigma^-)\mu = 0$ in which case $k = R(\Sigma)$ and $\delta = \mu'\Sigma^- \mu$.*

The required condition is satisfied for all μ if Σ is a reflexive inverse of Σ and is satisfied for all Σ if and only if $\mu \in \mathscr{M}(\Sigma)$. We note further that if $\mu \in \mathscr{M}(\Sigma)$ then $Y \in \mathscr{M}(\Sigma)$ with probability 1. In such a case with probability 1, $Y'\Sigma^- Y$ is invariant with respect to choice of Σ^-.

It has come to our notice after the Berkeley Symposium that Bhapkar [38] has obtained the result stated in the Corollary 7.1 to our Theorem 7.3. But the result of Theorem 7.3 is more general and that of the corollary is only a particular case.

Condition (i), $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$, of Theorem 7.3 seems to have been found first by Ogasawara and Takahashi [39].

8. Discriminant function in multivariate analysis

8.1. *Singular multivariate normal distribution.* The book *Linear Statistical Inference and its Applications* [10] develops a density free approach to study the distribution and inference problems associated with a multivariate normal distribution. The approach is more general than the usual one since it includes the study of the normal distribution with a singular covariance matrix which does not admit a density in the usual sense. The elegance of the density free approach was further demonstrated by Mitra [40]. However, in some problems, as in the construction of a discriminant function, it is useful to have an explicit expression for the density. The density function of a multivariate normal distribution, as it is usually written, involves the inverse of the variance-covariance (dispersion) matrix, which necessitates the assumption that the dispersion matrix is nonsingular. In this section we demonstrate how the g inverse is useful in defining the density function and in extending some of the results developed for the nonsingular case to the singular distribution.

Let Y be a $p \times 1$ vector random variable. In Rao [10], Y is defined to have a p variate normal distribution if $m'Y$ has a univariate normal distribution for every vector $m \in \mathscr{R}^p$. In such a case it is shown that the distribution is characterized by the parameters

$$\mu = E(Y),\ \Sigma = E[(Y - \mu)(Y - \mu)'], \tag{8.1}$$

called the mean vector and the dispersion matrix of Y, respectively; the symbol $N_p(\mu, \Sigma)$ is used to denote the p variate normal distribution. The distribution is said to be singular if $R(\Sigma) = \rho < p$ in which case ρ is called the rank of the distribution and we may use the symbol $N_p(\mu, \Sigma(\rho))$ to specify the rank in addition to the basic parameters.

Let N be $p \times (p - \rho)$ matrix of rank $p - \rho$ such that $N'N = I$, $N'\Sigma = 0$ and A be a $p \times \rho$ matrix of rank ρ such that $N'A = 0$ and $A'A = I$. By construction $(N : A)$ is an orthogonal matrix. We make the transformation

$$Z_1 = N'Y, \qquad Z_2 = A'Y. \tag{8.2}$$

Then

$$\begin{aligned} E(Z_1) &= N'\mu, & E(Z_2) &= A'\mu, \\ D(Z_1) &= N'\Sigma N = 0, & D(Z_2) &= A'\Sigma A. \end{aligned} \tag{8.3}$$

It follows that there exists a constant vector ζ such that

$$Z_1 = N'Y = N'\mu = \zeta, \tag{8.4}$$

with probability 1 and since $A'\Sigma A$ is nonsingular Z_2 has the ρ variate normal density

$$(2\pi)^{-\rho/2}|A'\Sigma A|^{-1/2} \exp\{-\tfrac{1}{2}(Z_2 - A'\mu)'(A'\Sigma A)^{-1}(Z_2 - A'\mu)\}. \tag{8.5}$$

We observe that

$$|A'\Sigma A| = \lambda_1 \cdots \lambda_\rho, \tag{8.6}$$

where $\lambda_1, \cdots, \lambda_\rho$ are the nonzero eigenroots of Σ and

$$(Z_2 - A'\mu)'(A'\Sigma A)^{-1}(Z_2 - A'\mu) = (Y - \mu)'\Sigma^-(Y - \mu) \tag{8.7}$$

where Σ^- is any g inverse of Σ. Thus the density of Y on the hyperplane $N'(Y - \mu) = 0$ or $N'Y = \zeta$ is defined by

$$\frac{(2\pi)^{-\rho/2}}{(\lambda_1 \cdots \lambda_\rho)^{1/2}} \exp\{-\tfrac{1}{2}(Y - \mu)'\Sigma^-(Y - \mu)\} \tag{8.8}$$

which is an explicit function of the vector Y and its associated parameters μ and Σ. The expression (8.8) was considered by Khatri [41] in deriving some distributions in the case of a singular normal distribution.

8.2. *Discriminant function.* The density function derived in (8.8) can be used in determining the discriminant function (ratio of likelihoods) for assigning an individual as a member of one of two populations to which it may belong.

Let Y be a $p \times 1$ vector of observations which has the distribution $N_p[\mu_1, \Sigma_1(\rho_1)]$ in the first population and $N_p(\mu_2, \Sigma_2(\rho_2))$ in the second population. We shall construct the discriminant function applicable to different situations.

Case 1. $\Sigma_1 = \Sigma_2 = \Sigma$, $R(\Sigma) = \rho < p$, $N'\mu_1 \neq N'\mu_2$.

The distribution of Y consists of two parts as shown in (8.4) and (8.5) of which (8.4) is the almost sure part. If $N'\mu_1 \neq N'\mu_2$,

$$N'Y = N'\mu_1 = \zeta_{11} \quad \text{with probability 1} \tag{8.9}$$

if Y comes from the first population, and

$$N'Y = N'\mu_2 = \zeta_{21} \quad \text{with probability 1} \tag{8.10}$$

if Y comes from the second population. Then the discriminant function is $N'Y$, and in fact it provides perfect discrimination. No use need be made of the other part (8.5) of the distribution of Y.

Case 2. $\Sigma_1 = \Sigma_2 = \Sigma$, $R(\Sigma) = \rho < p$, $N'\mu_1 = N'\mu_2$.

In this case $N'Y$ does not provide any discrimination and we have to consider the density (8.8). The log densities for the two populations are (apart from a constant)

$$-\tfrac{1}{2}\log(\lambda_1 \cdots \lambda_\rho) - \tfrac{1}{2}(Y - \mu_1)'\Sigma^-(Y - \mu_1), \tag{8.11}$$

and

$$-\tfrac{1}{2}\log(\lambda_1 \cdots \lambda_\rho) - \tfrac{1}{2}(Y - \mu_2)'\Sigma^-(Y - \mu_2). \tag{8.12}$$

Taking the difference and retaining only the portion depending on Y we obtain the discriminant function

$$\delta'\Sigma^- Y, \qquad \text{where} \quad \delta = \mu_1 - \mu_2 \tag{8.13}$$

which is of the same form as in the nonsingular case ($\delta'\Sigma^{-1}Y$). Now

$$V(\delta\Sigma^- Y) = \delta'\Sigma^-\delta \tag{8.14}$$

which is the analogue of Mahalanobis distance $D^2 (= \delta'\Sigma^{-1}\delta)$ in the singular case.

Case 3. $\Sigma_1 \neq \Sigma_2$, $\mathscr{M}(\Sigma_1) \neq \mathscr{M}(\Sigma_2)$.

The discrimination is perfect as in Case 1. Let N be a matrix of maximum rank such that $N'\Sigma_1 N = 0 = N'\Sigma_2 N$, and let A be a matrix of maximum rank such that $(N : A)'\Sigma_1 = 0$, and let B be a matrix of maximum rank such that $(N : B)'\Sigma_2 = 0$. Finally let C be such that $(N : A : B : C)$ is a $p \times p$ matrix of rank p. Consider the transformation

$$Z_1 = N'Y,\ Z_2 = A'Y,\ Z_3 = B'Y,\ Z_4 = C'Y \tag{8.15}$$

The distributions of these variables in the two populations are given in Table III.

TABLE III

DISTRIBUTION OF THE DISCRIMINANT VARIABLES

Case 3.

Population	Z_1	Z_2	Z_3	Z_4
1	$\zeta_{11} = N'\mu_1$ with prob. 1	$\zeta_{12} = A'\mu_1$ with prob. 1	$N(B'\mu_1, B'\Sigma_1 B)$	$N(C'\mu_1, C'\Sigma_1 C)$
2	$\zeta_{21} = N'\mu_2$ with prob. 1	$N(A'\mu_2, A'\Sigma_2 A)$	$\zeta_{22} = B'\mu_2$ with prob. 1	$N(C'\mu_2, C'\Sigma_2 C)$

It is seen that the variables Z_1, Z_2 and Z_3 provide perfect discrimination unless $\zeta_{11} = \zeta_{21}$, $A = 0$ and $B = 0$, which can happen only when $\mathcal{M}(\Sigma_1) = \mathcal{M}(\Sigma_2)$.

Case 4. $\Sigma_1 \neq \Sigma_2$, $\mathcal{M}(\Sigma_1) = \mathcal{M}(\Sigma_2)$.

Let N be as defined in Case 3 and consider $N'Y$ which is a constant for both the populations. If $N'\mu_1 \neq N'\mu_2$, then we have perfect discrimination. If $N'\mu_1 = N'\mu_2$, then we have to consider the densities

$$(\lambda_1 \cdots \lambda_\rho)^{-1/2} \exp\left\{-\tfrac{1}{2}(Y - \mu_1)'\Sigma_1^-(Y - \mu_1)\right\} \tag{8.16}$$

and

$$(\lambda_1' \cdots \lambda_\rho')^{-1/2} \exp\left\{-\tfrac{1}{2}(Y - \mu_2)'\Sigma_2^-(Y - \mu_2)\right\}, \tag{8.17}$$

where $\lambda_1, \cdots, \lambda_\rho$ are the nonzero eigenvalues of Σ_1, $\lambda_1', \cdots, \lambda_\rho'$ are those of Σ_2 and Σ_1^-, Σ_2^- are any g inverses of Σ_1, Σ_2. Taking logarithm of the ratio of densities and retaining only the terms depending on Y we have the quadratic discriminant function

$$(Y - \mu_1)'\Sigma_1^-(Y - \mu_1) - (Y - \mu_2)'\Sigma_2^-(Y - \mu_2) \tag{8.18}$$

analogous to the expression in the nonsingular case.

REFERENCES

[1] E. H. MOORE, *General Analysis*, Philadelphia, American Philosophical Society, 1935.

[2] ———, "On the reciprocal of the general algebraic matrix" (abstract), *Bull. Amer. Math. Soc.*, Vol. 26 (1920), pp. 394–395.

[3] Y. Y. TSENG, "Generalized inverses of unbounded operators between two unitary spaces," *Dokl. Akad. Nauk. SSSR.*, Vol. 67 (1949), pp. 431–434.

[4] ———, "Properties and classifications of generalized inverses of closed operators," *Dokl. Akad. Nauk. SSSR*, Vol. 67 (1949), pp. 607–610.

[5] ———, "Virtual solutions and general inversions," *Uspehi. Mat. Nauk.*, Vol. 11 (1956), pp. 213–215.

[6] R. PENROSE, "A generalized inverse for matrices," *Proc. Cambridge Philos. Soc.*, Vol. 51 (1955), pp. 406–413.

[7] ———, "On best approximate solutions of linear matrix equations," *Proc. Cambridge Philos. Soc.*, Vol. 52 (1956), pp. 17-19.

[8] C. RADHAKRISHNA RAO, "Analysis of dispersion for multiply classified data with unequal numbers in cells," *Śankhyā*, Vol. 15 (1955), 253–280.

[9] ———, "A note on a generalized inverse of a matrix with applications to problems in mathematical statistics," *J. Roy. Statist. Soc. Ser. B*, Vol. 24 (1962) pp. 152–158.

[10] ———, *Linear Statistical Inference and its Applications*, New York, Wiley, 1965.

[11] ———, "On the theory of least squares when parameters are stochastic and its application to analysis of growth curves," *Biometrika*, Vol. 52 (1965), pp. 447–458.

[12] ———, "A study of large sample test criteria through properties of efficient estimates," *Śankhyā Ser. A*, Vol. 23 (1961), pp. 25–40.

[13] ———, "Generalized inverse for matrices and its applications in mathematical statistics," *Research papers in Statistics, Festschrift for J. Neyman*, New York, Wiley, 1966.

[14] ———, "Least squares theory using an estimated dispersion matrix and its application to measurement of signals," *Proceedings of the Fifth Berkeley Symposium on Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 1, pp. 355–372.

[15] T. N. E. GREVILLE, "The pseudo-inverse of a rectangular matrix and its application to the solution of systems of linear equations," *SIAM Rev.*, Vol. 1 (1959), pp. 38–43.

[16] A. BJERHAMMER, "Rectangular reciprocal matrices with special reference to geodetic calculations," *Bull. Géodésique*, Vol. 52 (1951). pp. 188–220.

[17] ———, "Application of the calculus of matrices to the method of least squares with special reference to geodetic calculations," *Kungl. Tekn. Högsk. Hand. Stockholm.* No. 49 (1951), pp. 1–86.

[18] ———, "A generalized matrix algebra," *Kungl. Tekn. Högsk. Handl. Stockholm.* No. 124 (1958), pp. 1–32.

[19] A. BEN-ISRAEL and A. CHARNES, "Contributions to the theory of generalized inverses," *SIAM J. Appl. Math.*, Vol. 11 (1963), pp. 667–699.

[20] J. S. CHIPMAN, "On least squares with insufficient observations," *J. Amer. Statist. Assoc.*, Vol. 59 (1964), pp. 1078–1111.

[21] ———, "Specification problems in regression analysis," *Theory and Application of Generalized Inverses and Matrices*, Symposium Proceedings, Texas Technological College. Mathematics Series No. 4 (1968), pp. 114–176.

[22] J. S. CHIPMAN and M. M. RAO, "Projections, generalized inverses and quadratic forms," *J. Math. Anal. Appl.*, Vol. 9 (1964), pp. 1–11.

[23] J. E. SCROGGS and P. L. ODELL, "An alternative definition of the pseudo-inverse of a matrix," *SIAM J. Appl. Math.*, Vol. 14 (1966), pp. 796–810.

[24] R. C. BOSE, "Analysis of Variance," unpublished lecture notes, University of North Carolina, 1959.

[25] R. BOTT and R. J. DUFFIN, "On the algebra of networks," *Trans. Amer. Math. Soc.*, Vol. 74 (1953), pp. 99–109.

[26] H. CHERNOFF, "Locally optimal designs for estimating parameters," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 586–602.

[27] C. RADHAKRISHA RAO, "Calculus of generalized inverse of matrices. Part I: General theory," *Śankhyā Ser. A*, Vol. 29 (1967), pp. 317–342.

[28] S. K. MITRA, "On a generalized inverse of a matrix and applications," *Śankhyā Ser. A*, Vol. 30 (1968), pp. 107–114.

[29] ———, "A new class of g-inverse of square matrices," *Śankhyā Ser. A*, Vol. 30 (1968), pp. 323–330.

[30] P. BHIMAŚANKARAM and S. K. MITRA, "On a theorem of Rao on g-inverses of matrices," *Śankhyā Ser. A*, Vol. 31 (1969) pp. 365–368.

[31] S. K. MITRA and P. BHIMAŚANKARAM, "Some results on idempotent matrices and a matrix equation connected with the distribution of quadratic forms," *Śankhyā Ser. A.*, Vol. 32 (1970), pp. 353–356.

[32] S. K. MITRA and C. RADHAKRISHNA RAO, "Simultaneous reduction of a pair of quadratic forms," *Śankhyā Ser. A*, Vol. 30 (1968), pp. 313–322.

[33] ———, "Some results in estimation and tests of hypotheses under the Gauss-Markov model," *Śankhyā Ser. A*, Vol. 30 (1968), pp. 281–290.

[34] ———, "Conditions for optimality and validity of simple least squares theory," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 1617–1624.

[35] C. R. RAO, "A note on a previous lemma in the theory of least squares and some further results," *Śankhyā Ser. A*, Vol. 30 (1968), pp. 245–252.

[36] J. S. FRAME, "Matrix operations and generalized inverses," *IEEE Spectrum*, March (1964), pp. 209–220.

[37] M. SIBUYA, "Subclasses of generalized inverses of matrices," *Ann. Inst. Statist., Math.*, in press.

[38] V. P. BHAPKAR, "The invariance and distribution of the quadratic form $x'\Sigma^{-}x$," Technical Report, University of Kentucky, Department of Statistics, 1970.

[39] T, OGASAWARA and M. TAKAHASHI, "Independence of quadratic forms in normal system," *Journal of Science of the Hiroshima University*, Vol. 15 (1951), pp. 1–9.

[40] S. K. MITRA, "A density-free approach to matrix variate beta distribution," *Śankhyā Ser. A*, Vol. 32 (1970), pp. 81–88.

[41] C. G. KHATRI, "Some results for the singular multivariate regression models," *Śankhyā, Ser. A*, Vol. 30 (1968), pp. 267–280.