A Note on a Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics

By C. Radhakrishna Rao
Indian Statistical Institute, Calcutta

[Received November 1961]

Summary

Some years ago the author defined a pseudo inverse of a singular matrix and used it in representing a solution of normal equations and for obtaining variances and covariances of estimates in the theory of least squares (Rao, 1955). This provided a unified approach to least squares theory, including the case when the normal equations become singular. This note attempts to collect a few mathematical results, some of which are known in literature, associated with the inversion of singular and rectangular matrices, and to indicate briefly their use in problems of mathematical statistics.

1. Introduction

Consider linear equations $A\mathbf{x} = \mathbf{y}$, where x and y are column vectors of order n and m and A is a matrix of order $m \times n$. When $m = n$ and $|A| \neq 0$, an explicit representation of the solution is, $\mathbf{x} = A^{-1}\mathbf{y}$, where A^{-1} is the inverse of A such that $A^{-1}A = AA^{-1} = I$. The corresponding problem in the general case where A may be a rectangular or a square singular matrix has also received some attention from time to time, and attempts have been made to define a generalized inverse with properties similar to those of an inverse of a non-singular matrix (Moore, 1935; Murray and von Neumann, 1936; Bjerhammer, 1951; Penrose, 1955). Such a problem arises in the theory of least squares when the matrix of normal equations is singular. The author gave (Rao, 1955) a method of computing what is called a pseudo inverse (p-inverse) of a singular matrix with the help of which an explicit representation of a solution to normal equations and expressions for variances of estimates could be obtained. The term pseudo inverse is also used in a paper by Greville (1957). Wilkinson (1958) uses the term "effective inverse" in such cases and mentions that the application to normal equations with a singular matrix was suggested to him by A. T. James in a personal communication in 1956.

In the present note the methods introduced in earlier papers are extended to answer other questions which arise in the solution of linear equations involving rectangular or singular matrices, such as their consistency and the uniqueness of certain linear functions. These problems are of some practical interest and do not seem to have been considered in any earlier work on this subject. In a recent paper (Rao, 1961) the author used a p-inverse in the derivation of distributions of quadratic forms when the variables have a singular multivariate normal distribution. Some further applications in this direction are also given in this note.

It is preferable to use the term "generalized inverse" (g-inverse) to cover both the regular and pseudo inverses and restrict the term p-inverse to cases where a true inverse does not exist.

2. Definition and Properties of g-inverse

Definition. A generalized inverse (or g-inverse) of a matrix A of order $m \times n$ is a matrix of order $n \times m$, denoted by A^-, such that for any $\mathbf{y}$ for which $A\mathbf{x} = \mathbf{y}$ is consistent, $\mathbf{x} = A^-\mathbf{y}$ is a solution.

An inverse so defined is not necessarily unique, but may be made unique by imposing further conditions. Some of the earlier writers considered the problem of defining a unique inverse and studying its properties. But this is not necessary for our purpose.

Lemma 1. If A^- is a g-inverse then $AA^-A = A$, and conversely.

Choose $\mathbf{y}$ as the ith column $\mathbf{a}_i$ of A. Then the equation $A\mathbf{x} = \mathbf{a}_i$ is obviously consistent and hence $\mathbf{x} = A^-\mathbf{a}_i$ is a solution, i.e. $AA^-\mathbf{a}_i = \mathbf{a}_i$ for all i, which implies that $AA^-A = A$.

Conversely if A^- exists such that $AA^-A = A$, and $A\mathbf{x} = \mathbf{y}$ is consistent, then $AA^-A\mathbf{x} = A\mathbf{x}$ or $AA^-\mathbf{y} = \mathbf{y}$. Hence $\mathbf{x} = A^-\mathbf{y}$ is a solution.

Lemma 2. Let A^- be a g-inverse of A of order $m \times n$ and B a matrix of order $n \times s$ admitting a right (regular) inverse B_r^{-1} of order $s \times n$. Then $B_r^{-1}A^-$ is a g-inverse of AB.

The result is true since

$$ABB_r^{-1}A^-AB = AIA^-AB = AA^-AB = AB.$$

We note that the result may not be true if no regular inverse exists for B and we substitute any g-inverse of B.

Lemma 3. A g-inverse exists for any matrix A, although it may not be unique, and further it can be constructed in such a way that it has A itself as a g-inverse. Or in other words it is possible to find A^- such that $AA^-A = A$ and $A^-AA^- = A^-$.

Given a matrix A of order $m \times n$, there exist non-singular matrices P and Q of orders m and n respectively such that

$$PAQ = \Delta \quad \text{or} \quad A = P^{-1}\Delta Q^{-1},$$

where

$$\Delta = \left(\begin{array}{c|c} D_s & 0 \\ \hline 0 & 0 \end{array}\right)$$

and D_s is a diagonal matrix of order s and rank s.

Let us define

$$A^- = Q\Delta^-P,$$

where

$$\Delta^- = \left(\begin{array}{c|c} D_s^{-1} & 0 \\ \hline 0 & 0 \end{array}\right).$$

Then it may be seen that

$$AA^-A = P^{-1}\Delta Q^{-1} = A.$$

Hence A^- is a g-inverse. Also

$$A^-AA^- = Q\Delta^-P = A^-.$$

In fact any matrix A admits the decomposition $P\Delta Q'$ where P and Q are orthogonal matrices, in which case A^- defined by $Q\Delta^-P'$ has the additional properties $(AA^-)' = AA^-$ and $(A^-A)' = A^-A$.

Moore (1935) and Penrose (1955) have shown that there exists a unique matrix B satisfying the equations $ABA = A$, $BAB = B$, $(AB)' = AB$, $(BA)' = BA$. Such a matrix B is defined by them to be the generalized inverse of A. In this paper a generalized inverse stands for any matrix satisfying the definition of section 2. We shall, however, show how certain special choices of a g-inverse satisfying conditions other than those specified by Moore and Penrose are useful in practical problems.

Lemma 4. Let $A^- A = H$ for a given g-inverse A^-. Then

(i) $H^2 = H$, i.e. H is idempotent;

(ii) $AH = A$;

(iii) the solutions of $A\mathbf{x} = 0$ can be expressed as $(H-I)\mathbf{z}$, where $\mathbf{z}$ is arbitrary;

(iv) a general solution of $A\mathbf{x} = \mathbf{y}$, when consistent, is $A^-\mathbf{y}+(H-I)\mathbf{z}$;

(v) $\mathbf{q}'\mathbf{x}$ has a unique value for all $\mathbf{x}$ satisfying the equation $A\mathbf{x} = \mathbf{y}$, if $\mathbf{q}'H = \mathbf{q}'$.

Since $AA^-A = A$ we have, multiplying both sides by A^-, that $H^2 = H$, which proves (i). Point (ii) follows by replacing A^-A by H in $AA^-A = A$.

Since $AH = A$, $\text{rank}(H) \geq \text{rank}(A)$. But $A^-AH = H$ and therefore

$$\text{rank}(H) \leq \text{rank}(A).$$

Hence
$$\text{rank}(H) = \text{rank}(A).$$

Also
$$\text{rank}\{(H-I)\} = n - \text{rank}(H) = n - \text{rank}(A),$$

where n is the number of columns in A. Since $A(H-I) = 0$, the columns of $(H-I)$ supply all the solutions of $A\mathbf{x} = 0$. Hence a general solution is $(H-I)\mathbf{z}$ where $\mathbf{z}$ is arbitrary. This proves (iii), and (iv) follows from (iii) since $A^-\mathbf{y}$ is a particular solution of $A\mathbf{x} = \mathbf{y}$. Result (v) follows by substituting in $q'\mathbf{x}$ a general solution of $A\mathbf{x} = \mathbf{y}$ and expressing the condition that it is unique.

Lemma 4 shows how, when a g-inverse is supplied, some of the problems associated with linear equations can be solved by first computing H. But there does not seem to be an easy method of deciding whether $A\mathbf{x} = \mathbf{y}$ is consistent. This can be done in a simple way if one has a g-inverse as defined in Lemma 5 or as computed in the illustrative example of section 3.

To avoid complications in further algebraic treatment let us, by adding rows or columns of zeros whichever is necessary, make A a square matrix. The necessary change in the linear equations is to extend $\mathbf{y}$ by a number of zeros if rows are added and to extend $\mathbf{x}$ by additional variables if columns are added. This makes no difference to the original equations. In actual computations such a change need not be made, but this is done here only to provide some convenience in theoretical discussions. We shall therefore consider, without loss of generality, the equation $A\mathbf{x} = \mathbf{y}$ where A is a square matrix, observing that the g-inverse appropriate to the original matrix can always be obtained from the g-inverse of the extended square matrix by omitting some rows or columns.

Let us first recall that given a matrix A, there exists a non-singular matrix B such that $BA = H$, where H has the following properties:

(i) The diagonal elements are zeros or unities.

(ii) When a diagonal element, say the ith, is unity all elements in the ith column and all elements preceding unity in the ith row are zero.

(iii) When a diagonal element is zero, say the jth, all elements in the jth row are zero, and also those elements below the zero diagonal element in the jth column.

This is exactly the form obtained by starting with any matrix and applying the method of sweep out and interchange of rows, if necessary, to bring unities to the diagonal. Each operation in the sweep out or interchange of rows is equivalent to premultiplication of the given matrix by a non-singular matrix. Let B be the product of all matrices involved in the step-by-step reduction of the original matrix. Define G as a diagonal matrix with its ith diagonal element unity if the ith diagonal element of H is zero, and zero otherwise.

Lemma 5. With A, B, H and G as defined above the following are true:

(i) H is idempotent;

(ii) $AH =$ ~~H~~; A

(iii) $ABA = A$ and hence B is a g-inverse;

(iv) The necessary and sufficient condition that $A\mathbf{x} = \mathbf{y}$ is consistent is $GB\mathbf{y} = 0$, i.e. if the r_1th, r_2th, ... rows in H are null, then the r_1th, r_2th, ... elements in $B\mathbf{y}$ must be zero;

(v) A general solution of $A\mathbf{x} = \mathbf{y}$ is $B\mathbf{y}+(H-I)\mathbf{z}$, where $\mathbf{z}$ is an arbitrary column vector;

(vi) The necessary and sufficient condition that a linear function $\mathbf{q}'\mathbf{x}$ is unique, when $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{y}$ is $\mathbf{q}'H = \mathbf{q}'$.

Point (i) is proved by verification and (ii) follows from

$$BA = H, \quad A = B^{-1}H, \quad AH = B^{-1}H^2 = B^{-1}H = A.$$

Point (iii) is proved by multiplying both sides of $BA = H$ by A and using the result $AH = A$. We note that B is a non-singular inverse, although A may be singular. Such a non-singular inverse is necessary for the consistency test given in point (iv). The conditions imposed by Murray and Penrose necessarily lead to a singular inverse when A is singular.

If the equation $A\mathbf{x} = \mathbf{y}$ is consistent, so is $BA\mathbf{x} = B\mathbf{y}$ or $H\mathbf{x} = B\mathbf{y}$ and conversely, since $|B| \neq 0$. If the rth row of H is zero, then the rth element of $H\mathbf{x}$ is zero and so must be the rth element of $B\mathbf{y}$. Conversely, if this is true, $\mathbf{x} = B\mathbf{y}$ is obviously a solution of $H\mathbf{x} = B\mathbf{y}$. This proves (iv), which is an important result. If in addition to B, we know which of the rows of H are null, we have an automatic test for consistency of $A\mathbf{x} = \mathbf{y}$ while finding a solution. Let the r_1th, r_2th, ..., r_kth rows in H be null. Then we need only compute $B\mathbf{y}$ and examine whether the r_1th, r_2th, ..., r_kth elements are zero. If they are, the equations are consistent, in which case $B\mathbf{y}$ itself is a solution. Points (v) and (vi) are proved as in Lemma 4.

The knowledge of B and H thus enables us to answer all questions connected with the linear equations $A\mathbf{x} = \mathbf{y}$, while B alone is sufficient to find a solution when $A\mathbf{x} = \mathbf{y}$ is consistent.

3. The Computation of g-inverse

A g-inverse considered in Lemma 5 is computed in exactly the same way as a regular inverse when it exists. We shall illustrate the technique by means of an example. Let A be as shown in Table 1, which also contains an appended unit matrix. The method of sweep out is applied to the entire matrix.

The computations may be abridged to a large extent by omitting some intermediate steps, but are presented in full for illustrating the method. Let us test whether for $\mathbf{y}' = (1, 1, 2, 1)$, $A\mathbf{x} = \mathbf{y}$ is soluble. Computing, we find $\mathbf{y}'B' = (\cdot 51, \cdot 2, \cdot 2, 0)$, which shows that the equation is soluble and $(\cdot 51, \cdot 2, \cdot 2, 0)$ is itself a solution. On the other hand, for $\mathbf{y}' = (0, 0, 1, 0)$, $\mathbf{y}'B' = (0, 0, 0, 1)$. Since the last element is not zero no solution exists. Let us examine whether $x_1+x_2+x_3+x_4$ has a unique value. The condition is $\mathbf{q}'H = \mathbf{q}'$, $\mathbf{q}$ being $(1, 1, 1, 1)$ in the present example. Verifying $\mathbf{q}'H = (1, 1, 1, 2) \neq (1, 1, 1, 1)$, we see that the given function of the variables is not uniquely determined.

TABLE 1

Computation of the matrices H and B

matrix A				*Unit matrix*			
·	*2*	3	1	1	·	·	·
1	1	1	2	·	1	·	·
1	3	4	3	·	·	1	·
·	·	5	1	·	·	·	1
·	1	1·5	·5	·5	·	·	·
1	·	−·5	1·5	−·5	1	·	·
1	·	−·5	1·5	−1·5	·	1	·
·	·	5	1	·	·	·	1
·	1	1·5	·5	·5	·	·	·
1	·	−·5	1·5	−·5	1	·	·
·	·	·	·	−1	−1	1	·
·	·	*5*	1	·	·	·	1
·	1	·	·2	·5	·	·	−·3
1	·	·	1·6	−·5	1	·	·01
·	·	·	·	−1	−1	1	·
·	·	1	·2	·	·	·	·2
1	·	·	1·6	−·5	1	·	·01
·	1	·	·2	·5	·	·	−·3
·	·	1	·2	·	·	·	·2
·	·	·	·	−1	−1	1	·
matrix H				*matrix B*			

The second block is obtained by taking 2 (in italics) as a pivot and then, by successive subtraction, sweeping out the second column. The third block is obtained by taking 1 (in italics) as a pivot and sweeping out the first column. The fourth block is obtained by choosing 5 (in italics) as a pivot and sweeping out the third column. The final block, obtained simply by interchanging of rows, gives the matrices H and B.

4. APPLICATION TO LEAST SQUARES THEORY

In the theory of least squares, we have a vector of n observations such that

$$E(\mathbf{y}) = A\boldsymbol{\tau}, \quad D(\mathbf{y}) = \sigma^2 I,$$

where A is $n \times m$ matrix with known coefficients and $\boldsymbol{\tau}$ is $m \times 1$ vector of unknown parameters. According to the theory of least squares we minimize $(\mathbf{y} - A\boldsymbol{\tau})'(\mathbf{y} - A\boldsymbol{\tau})$ and obtain the normal equations, $A'A\mathbf{t} = A'\mathbf{y}$, which can be shown to be consistent for any $\mathbf{y}$. Let C be any g-inverse of $A'A$. Then $\mathbf{t} = CA'\mathbf{y}$ is a solution of the normal equations. If $|A'A| = 0$, the solution is not unique, but, however, a linear function $\mathbf{p}'\boldsymbol{\tau}$ may be unique. Such a function is said to be estimable and its estimate is given by $\mathbf{p}'\mathbf{t} = \mathbf{p}'CA'\mathbf{y}$. We have the following results.

(i) $E(\mathbf{p}'\mathbf{t}) = \mathbf{p}'\boldsymbol{\tau}$ if $\mathbf{p}'\boldsymbol{\tau}$ is estimable;

(ii) $V(\mathbf{p}'\mathbf{t}) = \mathbf{p}'C\mathbf{p}\sigma^2$;

(iii) $V(\mathbf{p}'\mathbf{t}) \leqslant V(\mathbf{b}'\mathbf{y})$ for any $\mathbf{b}$ such that $E(\mathbf{b}'\mathbf{y}) = \mathbf{p}'\boldsymbol{\tau}$ or $\mathbf{b}'A = \mathbf{p}'$.

Since p′t is unique we have, by Lemma 2, $\mathbf{p}'CA'A = \mathbf{p}'$. Thus

$$E(\mathbf{p}'\mathbf{t}) = E(\mathbf{p}'CA'\mathbf{y}) = \mathbf{p}'CA'A\boldsymbol{\tau} = \mathbf{p}'\boldsymbol{\tau}$$

which proves (i). Point (ii) is established by straightforward computation

$$V(\mathbf{p}'\mathbf{t}) = V(\mathbf{p}'CA'\mathbf{y}) = \mathbf{p}'CA'AC'\mathbf{p}\sigma^2 = \mathbf{p}'C\mathbf{p}\sigma^2.$$

The proof of (iii) follows from the relation

$$\begin{aligned}\operatorname{cov}((\mathbf{p}'\mathbf{t}-\mathbf{b}'\mathbf{y}), \mathbf{p}'\mathbf{t}) &= (\mathbf{p}'CA'-\mathbf{b}')AC'\mathbf{p}\sigma^2\\ &= (\mathbf{p}'C\mathbf{p}-\mathbf{p}'C\mathbf{p})\sigma^2\\ &= 0.\end{aligned}$$

These results show that t and $\sigma^2 C$ may be regarded as estimates of $\boldsymbol{\tau}$ and the dispersion matrix of estimates respectively, for purposes of building up an estimate of any estimable function $\mathbf{p}'\boldsymbol{\tau}$ and determining its variance. The matrix C can be any g-inverse and not necessarily as defined in Lemma 5 or as computed in section 3 for certain other purposes. We thus have a unified treatment of least squares theory which is still considered in the literature under two separate cases depending on the singularity or non-singularity of $A'A$. The expression for the least sum of squares is $\mathbf{y}'\mathbf{y}-\mathbf{y}'A\mathbf{t} = \mathbf{y}'\mathbf{y}-\mathbf{y}'CAA'\mathbf{y}$.

It may be noted that in order to prove the result (i), estimability of $\mathbf{p}'\boldsymbol{\tau}$ may be defined in an alternative way, viz., by the existence of a linear function of y with expectation equal to $\mathbf{p}'\boldsymbol{\tau}$. Further, the same kind of proof is applicable when either the variables are correlated or there are restrictions on the parameters.

5. Singular Multivariate Normal Distribution

In a recent paper (Rao, 1961), the author considered distributions of quadratic forms of variables whose limiting distribution is multivariate normal with a singular dispersion matrix. The results of Lemmas 2b and 6 in the paper cited are of special interest. We shall give here a few other results to demonstrate the use of a p-inverse.

Let us consider a vector of p variables x with $E(\mathbf{x}) = 0$ and $D(\mathbf{x}) = G$, the dispersion matrix, of rank $k \leqslant p$.

Lemma 6. The necessary and sufficient condition that a quadratic form $\mathbf{x}'A\mathbf{x}$ has a χ^2 distribution is that G is a g-inverse of A.

The result is well known when $|G| \neq 0$. In any case G can be written

$$G = C\Delta C',$$

where C is an orthogonal matrix and Δ is a diagonal matrix with non-negative elements. Consider the transformation $\mathbf{y} = C\mathbf{x}$. Then y is normally distributed with $E(\mathbf{y}) = 0$ and

$$D(\mathbf{y}) = C'GC = \Delta.$$

The quadratic form $\mathbf{x}'A\mathbf{x}$ transforms to $\mathbf{y}'F\mathbf{y}$ where $F = C'AC$. In terms of the new variables in y, which are independently distributed, the condition that $\mathbf{y}'F\mathbf{y}$ has a χ^2-distribution is obviously $F\Delta F = F$. Writing in terms of A and C, we have

$$C'AC\Delta C'AC = C'AGAC = C'AC.$$

The last equality implies that $AGA = A$, which proves the required result. The χ^2 distribution has degrees of freedom equal to rank GA.

Consider the particular quadratic form $\mathbf{x}'G^-\mathbf{x}$ where $G^- = C\Delta^- C'$, and Δ^- is obtained from Δ by replacing the non-zero elements by their reciprocals. Applying the test of Lemma 6, we find

$$G^- G G^- = C\Delta^- C' C\Delta C' C\Delta^- C'$$
$$= C\Delta^- C' = G^-.$$

Hence $\mathbf{x}'G^-\mathbf{x}$ has a χ^2-distribution with degrees of freedom equal to k, the rank of G.

REFERENCES

BJERHAMMAR, A. (1951), "Rectangular reciprocal matrices, with special reference to geodetic calculations", *Bull. geod. Int.*, 52, 188–220.

GREVILLE, T. N. E. (1957), "The pseudo inverse of a rectangular or singular matrix and its application to the solution of systems of linear equations", *S.I.A.M. Newsletter*, 5, 3–6.

MOORE, E. H. (1935), *General Analysis*. Philadelphia: American Phil. Soc.

MURRAY, F. J. and VON NEUMANN, J. (1936), "On rings of operators", *Ann. Math. Princeton*, 37, 141–143.

PENROSE, R. A. (1955), "A generalized inverse for matrices", *Proc. Camb. phil. Soc.*, 51, 406–413.

RAO, C. R. (1955), "Analysis of dispersion for multiply classified data with unequal numbers in cells", *Sankhyā*, 15, 253–280.

—— (1961), "A study of large sample test criteria through properties of efficient estimates", *Sankhyā*, 23, 25–40.

WILKINSON, G. N. (1958), "Estimation of missing values for the analysis of incomplete data", *Biometrics*, 14, 257–286.