



STATE UNIVERSITY OF NEW YORK AT STONY BROOK

COLLEGE OF
ENGINEERING

REPORT No. 70

A DIRECT METHOD FOR GENERALIZED MATRIX INVERSION

by

R. P. Tewarson

AUGUST 15, 1965

Spec

TAI

N532

NO. 70

6-2

A DIRECT METHOD FOR GENERALIZED MATRIX INVERSION

R. P. Tewarson*

Abstract. A method for computing the generalized inverse of a matrix is described. It uses the well-known Gauss-Jordan elimination scheme in conjunction with the conventional Gramm-Schmidt orthogonalization process.

1. Introduction. Moore [11], Bjerhammar [6] and Penrose [14] have independently generalized the concept of matrix inversion. This paper is an addition to the already growing literature on the computational aspects of generalized inverses [2-5, 7, 9, 12, 15, 18, 19, 21]. As in [21] we shall consider a $m \times n$ matrix G of rank r with real coefficients. Since $[G^T]^+ = [G^+]^T$, (where T denotes the transpose and $+$ the generalized inverse of G), there is no loss of generality in assuming $m \leq n$.

2. Main Results. Penrose [13] states that for any matrix G there exist elementary permutation matrices H and K such that

$$(1) \quad HGK = A = \begin{bmatrix} N & B \\ C & CN^{-1}B \end{bmatrix},$$

where N is a square non-singular matrix of the same rank as A . Evidently N is $r \times r$, C is $(m-r) \times r$, B is $r \times (n-r)$ and $CN^{-1}B$ is $(m-r) \times (n-r)$. For the elementary permutation matrices H and K we have $H^{-1} = H^T$ and $K^{-1} = K^T$, therefore $G = H^T A K^T$. Hence from theorem II, Equation (12) of [8] we have

$$\begin{aligned} G^+ &= [K^T]^{-1} A^+ [H^T]^{-1} \\ &= K A^+ H. \end{aligned}$$

* This work was supported in part by the National Aeronautics and Space Administration, Washington, D. C., Grant No. NGR-33-015-013.

Therefore, in order to evaluate G^+ , it is desirable to have a scheme to evaluate A^+ as well as the elementary permutation matrices H and K . Making use of (1) it is easily verified that

$$2) \quad A = \begin{bmatrix} N \\ C \end{bmatrix} [I, \Delta],$$

where I is a unit matrix of order r and $\Delta = N^{-1}B$. Since N is a non-singular matrix, the columns of the matrix $\begin{bmatrix} N \\ C \end{bmatrix}$ are linearly independent; also it is clear that $[I, \Delta]$ has linearly independent rows. Therefore, (as shown by Reville [10]), it follows that

$$3) \quad A^+ = [I, \Delta]^+ \begin{bmatrix} N \\ C \end{bmatrix}^+,$$

where

$$4) \quad [I, \Delta]^+ = \begin{bmatrix} I \\ \Delta^T \end{bmatrix} \left\{ [I, \Delta] \begin{bmatrix} I \\ \Delta^T \end{bmatrix} \right\}^{-1} \\ = \begin{bmatrix} I \\ \Delta^T \end{bmatrix} [I + \Delta \Delta^T]^{-1},$$

and

$$\begin{aligned} \begin{bmatrix} N \\ C \end{bmatrix}^+ &= \left\{ [N^T, C^T] \begin{bmatrix} N \\ C \end{bmatrix} \right\}^{-1} [N^T, C^T] \\ &= [N^T N + C^T C]^{-1} [N^T, C^T] \\ &= \left[N^T \left\{ I + [N^T]^{-1} C^T C N^{-1} \right\} N \right]^{-1} [N^T, C^T] \\ &= N^{-1} \left\{ I + [C N^{-1}]^T [C N^{-1}] \right\}^{-1} [N^T]^{-1} [N^T, C^T], \end{aligned}$$

or

$$5) \quad \begin{bmatrix} N \\ C \end{bmatrix}^+ = N^{-1} \left\{ I + [C N^{-1}]^T [C N^{-1}] \right\}^{-1} \begin{bmatrix} I \\ [C N^{-1}]^T \end{bmatrix}.$$

The Gramm-Schmidt orthogonalization process can now be used for the determination of the values of $[I + \Delta\Delta^T]^{-1}$ in (4) and $\{I + [CN^{-1}]^T[CN^{-1}]\}^{-1}$ in (5). In order to do so the following result, due to Rust, Burrus and Schneberger [19] as stated in their equations (30) and (31), will be used.

Lemma. If the Gramm-Schmidt orthogonalization is performed on the columns of matrix $\begin{bmatrix} U \\ I \end{bmatrix}$, where U is any given $s \times t$ matrix with real coefficients and I is an identity matrix of rank t, and if the resulting matrix is denoted by $\begin{bmatrix} Q \\ P \end{bmatrix}$, where P is $t \times t$, then

$$(6) \quad [I + U^T U]^{-1} = P P^T$$

and

$$(7) \quad [I + U U^T]^{-1} = [I - Q Q^T].$$

We are now in a position to prove the following theorems:

Theorem 1. If the Gramm-Schmidt orthogonalization is performed on the columns of the matrix $\begin{bmatrix} CN^{-1} \\ I \end{bmatrix}$ to give $\begin{bmatrix} \hat{Q} \\ P \end{bmatrix}$ then

$$(8) \quad \{I + [CN^{-1}]^T[CN^{-1}]\}^{-1} = P P^T,$$

where I and P are $r \times r$. However, if the orthogonalization is performed on $\begin{bmatrix} [CN^{-1}]^T \\ I \end{bmatrix}$ to give $\begin{bmatrix} Q \\ P \end{bmatrix}$, then

$$(9) \quad \{I + [CN^{-1}]^T[CN^{-1}]\}^{-1} = I - Q Q^T,$$

where I is $(m-r) \times (m-r)$ and Q is $r \times (m-r)$.

Proof. If in the Lemma U is replaced by CN^{-1} , then (6) leads to (8).

Similarly, if $[CN^{-1}]^T$ replaces U in (7), then (9) is obtained, since $\{[CN^{-1}]^T\}^T = CN^{-1}$.

Theorem 2. If the Gramm-Schmidt orthogonalization process is performed on the columns of the matrix $\begin{bmatrix} \Delta \\ I \end{bmatrix}$ to give $\begin{bmatrix} S \\ P \end{bmatrix}$, then

$$(10) \quad [I + \Delta\Delta^T]^{-1} = [I - SS^T],$$

where I is $(n - r) \times (n - r)$ and S is $r \times (n - r)$. However, if the above orthogonalization is performed on $\begin{bmatrix} \Delta^T \\ I \end{bmatrix}$ to give $\begin{bmatrix} \hat{S} \\ R \end{bmatrix}$, then

$$(11) \quad [I + \Delta\Delta^T]^{-1} = RR^T,$$

where I and R are both $r \times r$.

Proof: Same as that of Theorem 1.

3. Computational Aspects. Suppose the Gauss-Jordan Elimination [17] is performed on the matrix A such that N is reduced to the identity matrix of order r . The elimination is equivalent to the premultiplication of A by a non-singular matrix E as follows:

$$(12) \quad EA = \begin{bmatrix} N^{-1} & 0 \\ -CN^{-1} & I \end{bmatrix} \begin{bmatrix} N & B \\ C & CN^{-1}B \end{bmatrix} \\ = \begin{bmatrix} I & \Delta \\ 0 & 0 \end{bmatrix}.$$

As is customary, the computation can be arranged such that at the completion of the elimination process the matrix A gets transformed as shown below:

$$(13) \quad A \longrightarrow \begin{bmatrix} N^{-1} & \Delta \\ -CN^{-1} & 0 \end{bmatrix}.$$

In practice the elimination is performed on G instead of A and therefore some other set of $m - r$ rows is reduced to zero rather than the last $m - r$ rows as shown in (12). Also due to the permutation matrix K , see (1), the columns of N^{-1} and Δ may not be positioned as indicated in (13). However, it is easy to

see that the position of all the pivots at the end of the elimination process on G determines the permutations H and K of (1).

In floating point computations, it is generally not easy to determine if some number is effectively zero or not. This fact leads to the following difficulty in the elimination process (12), namely, the problem of deciding whether a row of A has been transformed to zero or not. A technique, essentially due to Osborne [12, p. 304], will now be described for the above problem. Let $L_i^{(k)}$ denote the i th row of A after $k - 1$ pivots have been chosen, where $i = 1, 2, \dots, m$ and $k = 1, 2, \dots, r$. Since $L_i^{(k)}$ consists of multiples of the rows of A added to $L_i^{(1)}$, it is therefore reasonable to compare the Euclidean norm of $L_i^{(k)}$ to that of $L_i^{(1)}$ and use $\|L_i^{(k)}\|/\|L_i^{(1)}\|$ as the criterion to decide whether a row is zero or not. In passing, we may also mention the following technique.

In the solution of large linear programming problems the redundant constraints (artificial vectors at zero level) and the contradictory constraints (artificial vectors at non-zero level - problem infeasible) are determined in the following manner. If no pivot greater than a certain "pivot tolerance" can be found in a certain row, then that row is considered to be a linear combination of the other rows in which pivots have already been chosen. (In connection with the design and writing of the Linear Programming Compiler ALPS [1] about one hundred Linear Programming problems were solved. These actual production problems were collected from diverse users of Linear Programming and ranged in size from 20 to 805 rows. A "pivot tolerance" of 10^{-5} appeared to be adequate for a 40 bit mantissa.) So much for the elimination process.

As in Theorem 1, the Gram-Schmidt orthogonalization process can now be used as follows. A suitable matrix I is appended to either CN^{-1} or its transpose in order to be able to make use of either (8) or (9) in (5). Evidently,

In order to save storage space and in general computational work, the equation chosen out of (8) and (9) is the one which requires I of the smaller dimension to be appended to the relevant matrix. In other words, if $m - r \geq r$ viz. $r \leq m/2$, then (8) is chosen, however if $r \geq m/2$, then (9) is chosen. Similarly, from Theorem 2, it follows that for $r \leq n - r$ viz. $r \leq n/2$, (11) should be used; on the other hand, if $r > n/2$, then the choice falls on (10). Since $\max r = m$, it should be noted that $r \leq n/2$ implies that $m \leq n/2$ but $r > n/2$ implies that $m > n/2$. The results of all of the above mentioned cases can now be stated in the form of the following theorem.

Theorem 3. The generalized inverse of A is given by:

$$(14a) \quad A^+ = \begin{bmatrix} I \\ \Delta^T \end{bmatrix} RR^T N^{-1} PP^T [I, [CN^{-1}]^T], \text{ if } r \leq m/2$$

$$(14b) \quad A^+ = \begin{bmatrix} I \\ \Delta^T \end{bmatrix} RR^T N^{-1} [I - QQ^T] [I, [CN^{-1}]^T], \text{ if } m/2 < r \leq n/2$$

$$(14c) \quad A^+ = \begin{bmatrix} I \\ \Delta^T \end{bmatrix} [I - SS^T] N^{-1} [I - QQ^T] [I, [CN^{-1}]^T], \text{ if } r > n/2.$$

Moreover, the maximum storage spaces required to compute A^+ by the above methods (denoted by S_1 , S_2 and S_3 respectively) are given by

$$S_1 = m(n + r),$$

$$S_2 = \max \{m(n + r), r(m + 2n - r)\},$$

$$S_3 = r(m - r) + n^2, \text{ if } r \leq \sqrt{n(n - m)} ; \\ = m(n + r), \text{ if } r > \sqrt{n(n - m)} .$$

Proof: If $r \leq m/2 \leq n/2$, then using (11) in (4), (8) in (5) and the results thus obtained in (3), we have (14a). If $m/2 < r \leq n/2$, then from (3), (4), (5), (9) and (11) we have (14b). To get (14c), we use (3), (4), (5), (9) and (10). If the usual programming techniques for saving storage are used for computing the product of matrices etc., then routine calculations regarding the storage spaces yield the values for S_1 , S_2 and S_3 . Note that a maximum of

m additional cells of storage are needed which have not been included in the values of S_1 , S_2 or S_3 given above (working storage).

4. An example. Consider the following matrix from [5]:

$$G = \begin{bmatrix} \textcircled{-1} & -1 & 0 & 0 & 1 & 1 \\ 0 & \textcircled{1} & -1 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 2 & -1 & 3 & -3 & 1 & -2 \end{bmatrix}$$

If the elimination is performed by choosing the circled elements of the above matrix as pivots (viz., H and K are identity matrices), then as in (13)

we have

$$\begin{bmatrix} N^{-1} & \Delta \\ -CN^{-1} & 0 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence

$$N^{-1} = \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}, \quad -CN^{-1} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}, \quad \Delta = \begin{bmatrix} 1 & -1 & 0 & -1 \\ -1 & 1 & -1 & 0 \end{bmatrix}$$

Since $r = 2$ and $m = 4$, viz. $r = m/2$, therefore (14a) should be used. Now

$$\begin{bmatrix} CN^{-1} \\ I \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{\text{G.S.}} \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{102} \\ 2/\sqrt{6} & 4/\sqrt{102} \\ 1/\sqrt{6} & -7/\sqrt{102} \\ 0 & 6/\sqrt{102} \end{bmatrix} = \begin{bmatrix} \hat{Q} \\ P \end{bmatrix}$$

Hence

$$P = \begin{bmatrix} 1/\sqrt{6} & -7/\sqrt{102} \\ 0 & 6/\sqrt{102} \end{bmatrix} \text{ and } PP^T = \frac{1}{102} \begin{bmatrix} 66 & -42 \\ -42 & 36 \end{bmatrix}$$

Also

$$\begin{bmatrix} \Delta^T \\ I \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 0 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{\text{G.S.}} N^{-1} = \begin{bmatrix} 1 & -1/\sqrt{3} \\ -1 & 1/\sqrt{3} \\ 0 & -2/\sqrt{3} \\ -1 & -1/\sqrt{3} \\ 1 & 1/\sqrt{3} \\ 0 & 2/\sqrt{3} \end{bmatrix} = \begin{bmatrix} \hat{A} \\ S \\ R \end{bmatrix}.$$

Hence

$$R = \frac{1}{2} \begin{bmatrix} 1 & 1/\sqrt{3} \\ 0 & 2/\sqrt{3} \end{bmatrix} \quad \text{and} \quad RR^T = \frac{1}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Finally, substituting the values of Δ , RR^T , N^{-1} , PP^T and CN^{-1} in (14a)

and performing the indicated multiplications, one gets

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

5. Concluding Remarks. The methods for computing A^+ , given by Rust, Burrus and Schneeberger [19] and by Ben-Israel and Wersan [5] generally require $m^2 + mn$ cells of storage; the method given by Osborne [12] requires more than $m^2 + mn$ cells. The storage space (S_1 , S_2 or S_3) required for the method described in this paper is generally less than that in the above mentioned methods since usually $r < m$. However, if N is a poorly conditioned matrix, then more sophisticated methods, often requiring more work and storage space, are recommended, e.g. [9, 15]. In any case it seems reasonable to assume that in a

significant number of cases the methods of this paper would have less difficulties during the course of the elimination process than [5], because NN^T is more "ill-conditioned" than N [20]. A Fortran IV program based on the method described in this paper is being written.

State University of New York
Stony Brook, N. Y.

1. ALPS. Honeywell E.D.P., DSI-275, Wellesley Hills, Mass., 1964.
2. A. BEN-ISRAEL, On iterative computation of generalized inverses and associated projections, to appear in J. SIAM, Ser. B.
3. _____, A note on an iterative method for generalized inversion of matrices, Math. of Comp., v. 20, 1966, pp. 439-440.
4. _____, An iterative method for computing the generalized inverse of an arbitrary matrix, Math. of Comp., v. 19, pp. 452-455.
5. A. BEN-ISRAEL AND S. J. WERSAN, An elimination method for computing the generalized inverse of an arbitrary complex matrix, J. ACM 10, v. 4, 1963, pp. 532-537.
6. A. BJERHAMMAR, Rectangular reciprocal matrices with special reference to geodetic calculations, Bull. Geodesique 1951, pp. 188-220.
7. J. C. G. BOOT, The computation of the generalized inverse of singular or rectangular matrices, Amer. Math. Mon., v. 70, pp. 302-303.
8. H. DECELL, An application of generalized matrix inversion to sequential least square parameter estimation, NASA TND-2830, 1965, Washington, D. C.
9. G. GOLUB AND W. KAHAN, Calculating the singular values and pseudo-inverse of a matrix, J. SIAM: Ser. B, v. 2, 1965, pp. 205-224.
10. T. N. E. GREVILLE, Some applications of the pseudoinverse of a matrix, SIAM Rev., v. 2, 1960, pp. 15-22.
11. E. H. MOORE, On the reciprocal of the general algebraic matrix, Abstract, Bull. Amer. Math. Soc., v. 26, 1920, pp. 394-395.
12. E. E. OSBORNE, Smallest least square solutions of linear equations, J. SIAM: Ser. B, v. 2, 1965, pp. 300-307.
13. R. PENROSE, On best approximate solutions of linear matrix equations, Proc. Cambridge Philos. Soc., v. 52, 1956, pp. 17-19.
14. _____, A generalized inverse for matrices. Ibid. v. 51, 1955, pp. 406-413.
15. V. PEREYRA AND J. B. ROSEN, Computation of the pseudo-inverse of a matrix of unknown rank, Tech. Rep. CSL3, Computer Science Division, Stanford University, 1964.
16. L. PYLE, Generalized inverse computations using the gradient projection method, J. ACM 11, v. 4, 1964, pp. 422-428.
17. A. RALSTON, A first course in numerical analysis, McGraw-Hill, New York, N. Y. 1965, pp. 399-401.

18. J. B. ROSEN, Minimum and basic solutions to singular linear systems, J. SIAM, v. 12, 1964, pp. 156-162.
19. G. RUST, W. R. BURRUS, AND C. SCHNEEBERGER, A simple algorithm for computing the generalized inverse of a matrix, Comm. ACM, v. 9, 1966, pp. 381-387.
20. O. TAUSSKY, Note on the condition of matrices, MTAC, v. 4, 1950, pp. 111-112.
21. R. P. TEWARSON, An elimination method for generalized matrix inversion, submitted for publication to Comm. ACM.