g-inverses are indeed an integral part of linear algebra and should be treated as such even at an elementary level.

Various techniques have been employed to deal with the problem of singularity. In the linear model context one may reparametrize the model to make it a 'full rank' model. A more recent approach which is theoretically quite appealing is to extend the notion of inverse to arbitrary matrices. This can be achieved in several ways but the following definition has survived the test of elegance and applicability. Let A be an $m \times n$ matrix. We say that an $n \times m$ matrix G is a generalized inverse (or simply, a g-inverse) of A if $AGA = A$. Clearly if A is square and nonsingular then A^{-1} is the only g-inverse it has. But in general a matrix has infinitely many g-inverses. This abundance of g-inverses results in a very rich theory.

A matrix over an arbitrary field admits a g-inverse. This fact is well-known and not difficult to prove. However, it is interesting to see the various approaches by which the statement can be established. It becomes apparent during the process that g-inverses are indeed an integral part of linear algebra and should be treated as such even at an elementary level.

We summarize several ways in which the existence of a g-inverse over the complex numbers can be established. Some of the arguments remain valid over more general fields. Also, with some extra effort one can derive the existence of more specialized g-inverses such as the least-squares and the minimum-norm g-inverse and the Moore–Penrose inverse. The emphasis is on bringing together several topics encountered in a standard linear algebra course.

We now introduce some definitions. We deal with complex matrices. The set of $m \times n$ complex matrices will be denoted by $C^{m \times n}$. The transpose of A is denoted A^T whereas A^* is the conjugate transpose. In the sequel we only give the definitions that are required in the text but do not offer any motivation or examples. For a leisurely development of these concepts one may consult the references mentioned in the last section.

For an $m \times n$ matrix A consider the usual Penrose equations

(1) $AGA = A$,
(2) $GAG = G$,
(3) $(AG)^* = AG$,
(4) $(GA)^* = GA$.

Recall that the $n \times m$ matrix G is called a g-inverse of A if it satisfies (1). We call G a least-squares g-inverse if it satisfies (1),(3), a minimum-norm g-inverse if it satisfies (1),(4) and the Moore–Penrose inverse if it satisfies (1)-(4). Any matrix A admits a unique Moore–Penrose inverse, which we denote as A^+. If A is $n \times n$ then G is called the group inverse of A if it satisfies (1),(2) and $AG = GA$. The matrix A has group inverse, which is unique, if and only if $\text{Rank}(A) = \text{Rank}(A^2)$.

Proofs of Existence of g-inverse

First observe that if A is square and nonsingular, then A^{-1} is the only g-inverse of A. Also, if the matrix B has full column rank (i.e., the columns of B are linearly independent), then the left inverses of B are precisely its g-inverses. A similar remark applies to matrices of full row rank. We now turn to the existence of g-inverses of an arbitrary matrix.

1. A Proof using Rank Factorization

Any matrix admits a rank factorization. Thus if A is an $m \times n$ matrix of rank r, then there exist matrices B and C of rank r and of order $m \times r$ and $r \times n$, respectively such that $A = BC$. Clearly, B admits a left inverse, say B^- and C admits a right inverse, say C^-. Then $G = C^- B^-$ is a g-inverse of A since $AGA = BCC^- B^- BC = BC = A$. Note that if we choose B^- to be a least-squares g-inverse (in fact such a g-inverse must be B^+ since B has full column rank) then $G = C^- B^-$ is a least-squares g-inverse of A. Similarly choosing C^- to be a minimum-norm g-inverse (which must be C^+) we obtain a minimum-norm g-inverse of A. Combining these observations, $A^+ = C^+ B^+$ is the Moore–Penrose inverse of A.

We also remark that A has group inverse if and only if CB is nonsingular. In such a case, it is easily verified that $G = B(CB)^{-2}C$ is the group inverse of A.

2. A Proof using the Rank Canonical Form

If A is $m \times n$ of rank r, then there exist nonsingular matrices P and Q of order $m \times m$ and $n \times n$, respectively, such that

$$A = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q.$$

It can be verified that for any U, V, W of appropriate dimensions,

$$\begin{bmatrix} I_r & U \\ V & W \end{bmatrix}$$

is a g-inverse of

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Then

$$G = Q^{-1} \begin{bmatrix} I_r & U \\ V & W \end{bmatrix} P^{-1}$$

is a g-inverse of A.

This approach makes it evident that if A is not a square nonsingular matrix, then it admits infinitely many g-inverses, since each choice of U, V, W leads to a different g-inverse. We can also deduce the following useful fact: if A has rank r, and if $r \le k \le \min\{m, n\}$ then A admits a g-inverse of rank k. To see this, choose U and V to be zero and W to be a matrix of rank $k - r$. In particular, any square matrix admits a nonsingular g-inverse.

3. A Computational Approach

Let A be an $m \times n$ matrix of rank r. Then A has an $r \times r$ nonsingular submatrix. For convenience we assume that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is $r \times r$ and nonsingular. Since the rank of A is r, the first r columns of A form a basis for the column space and hence $A_{12} = A_{11}X$ and $A_{22} = A_{21}X$ for some matrix X. Let

$$G = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

where the zero blocks are chosen so as to make G an $n \times m$ matrix. Then a simple computation shows that

$$\begin{aligned} AGA &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}A_{11}^{-1}A_{11}X \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}X \end{bmatrix} \\ &= A. \end{aligned}$$

If an $r \times r$ nonsingular submatrix of A is not located in the top left corner then the above algorithm requires a small modification as follows. Locate an $r \times r$ nonsingular submatrix, say B, of A. In A, replace B by the transpose of B^{-1} and all the remaining entries by zero. Transpose the resulting matrix to get a g-inverse of A.

The algorithm prescribed above provides a convenient way to calculate the g-inverse of a matrix of a small order which might arise in a classroom example. This is the reason for calling the proof 'computational'. Of course when the matrix is large, one must use more sophisticated methods in order to locate the nonsingular $r \times r$ matrix and to invert it.

As an example, consider the matrix

$$A = \begin{bmatrix} 1 & 7 & 1 & 4 \\ 1 & 2 & 0 & 1 \\ 0 & 5 & 1 & 3 \end{bmatrix}$$

of rank 2. For illustration, we pick the nonsingular 2×2 submatrix of A formed by rows 2, 3 and columns 2, 4. Replacing the matrix by its inverse transpose and the other entries by zero produces the matrix

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

Transposing the above matrix results in a g-inverse of A,

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

4. A Proof by Induction

Let A be an $m \times n$ matrix of rank r. We assume, without loss of generality, that $r < n \leq m$. Then A must have a column, which is a linear combination of the rest. For convenience, take it to be the last column. Write A as an augmented matrix $A = [B, x]$ where x is the last column of A. Proceeding by induction on $m + n$, we may assume that B has a g-inverse, say B^-. Set

$$G = \begin{bmatrix} B^- \\ 0 \end{bmatrix}.$$

We claim that G is a g-inverse of A. Observe that

$$AGA = BB^-A = BB^-[B, x] = [B, BB^-x].$$

Since x is in the column space of B, $x = By$ for some y. Then $BB^-x = BB^-By = By = x$. It follows that $AGA = A$ and the claim is proved.

More generally, suppose A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix};$$

where A_{11} is of rank r. Then

$$G = \begin{bmatrix} A_{11}^{-} & 0 \\ 0 & 0 \end{bmatrix}$$

is a g-inverse of A for any choice of a g-inverse A_{11}^{-} of A_{11}. (Here the zero blocks in G are chosen to make it an $n \times m$ matrix.)

5. An Operator Theoretic Approach

Consider the map f_A defined from $C^{m \times n}$ to itself given by $f_A(B) = AB^*A$. We show that A is in the range of f_A. On $C^{m \times n}$ we have the usual inner product $< X, Y > = \operatorname{tr}(XY^*)$. Suppose Z is orthogonal to the range of f_A. Then $\operatorname{tr}(AB^*AZ^*) = 0$ for all $B \in C^{m \times n}$ and hence $\operatorname{tr}(B^*AZ^*A) = 0$ for all $B \in C^{m \times n}$. It follows that $AZ^*A = 0$. Thus $AZ^*AZ^* = 0$ and hence $(AZ^*)^2 = 0$. Thus all eigenvalues of AZ^* are zero and therefore $\operatorname{tr}(AZ^*) = 0$. Thus Z is orthogonal to A. It follows that any matrix which is orthogonal to the range of f_A is orthogonal to A and hence A is in the range of f_A. Thus there exists $H \in C^{m \times n}$ such that $f_A(H) = AH^*A = A$. But then $G = H^*$ is a g-inverse of A.

6. A Functional Approach

Subspaces S and T of a vector space V are said to be complementary if $V = S + T$ and $S \cap T = \{0\}$. If A is an $m \times n$ matrix then we denote by $N(A) \subset C^n$ the null space of A and by $R(A) \subset C^m$ the range space of A. Let S be a subspace of C^n complementary to $N(A)$ and let T be a subspace of C^m complementary to $R(A)$. The matrix A defines a linear map from C^n to C^m. Let A_1 be the restriction of the map to S. Then it is easily seen that A_1 is a one-to-one map from S onto $R(A)$. Thus

$A_1^{-1}x; x \in R(A)$ is well-defined. Define $G: C^m \to C^n$ as $Gx = A_1^{-1}x$ if $x \in R(A)$; $Gx = 0$ if $x \in T$ and extend linearly to C^m. ($x \in C^m$ can be uniquely written as $x = y + z$, where $y \in R(A)$; $z \in T$. Then set $Gx = Gy + Gz$.) We denote the matrix of G with respect to the standard basis by G itself. Then it is easy to check that $AGA = A$ and hence we have got a g-inverse of A.

We remark that if $S = R(A^*)$ and $T = S^\perp = N(A^*)$ then the g-inverse constructed above is the Moore–Penrose inverse. When do we get a least-squares g-inverse and a minimum-norm g-inverse?

7. A Proof using the Spectral Theorem

Suppose A is an $n \times n$ hermitian matrix. By the spectral theorem, there exists a unitary matrix U such that

$$A = U \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} U^*,$$

where D is the diagonal matrix with the nonzero eigenvalues of A along its diagonal. Then

$$G = U^* \begin{bmatrix} D^{-1} & X \\ Y & Z \end{bmatrix} U$$

is a g-inverse of A for arbitrary X, Y, Z. Thus any hermitian matrix admits a g-inverse. Now for an arbitrary $m \times n$ matrix A, it can be verified that $G = (A^*A)^- A^*$ is a g-inverse of A for an arbitrary g-inverse $(A^*A)^-$ of the hermitian matrix A^*A.

8. A Proof using the Singular Value Decomposition

If A is an $m \times n$ matrix then there exist unitary matrices U and V such that

$$A = U \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} V,$$

where D is the diagonal matrix with the singular values of A along its diagonal. This is the well-known singular value decomposition of A. Then

$$G = V^*\begin{bmatrix} D^{-1} & X \\ Y & Z \end{bmatrix}U^*$$

is a g-inverse of A for arbitrary X, Y, Z. If $X = 0$ then G is a least-squares g-inverse, if $Y = 0$ then G is a minimum-norm g-inverse and finally, if X, Y and Z are all null matrices then G is the Moore–Penrose inverse.

9. A Proof using Linear Equations

If A is an $m \times n$ matrix then recall that the column space of A^* equals the column space of A^*A. Therefore the equation $A^*AX = A^*$ is consistent. Thus there exists an $n \times m$ matrix G such that $A^*AG = A^*$. Now

$$AGA = G^*A^*AGA = G^*A^*A = A$$

and hence G is a g-inverse of A.

10. A Proof Employing a Bordered Matrix

Let A be an $m \times n$ matrix. Let S be a subspace of C^m complementary to $R(A)$ and let T be a subspace of C^n complementary to $R(A^T)$. Let X be an $m \times (m - r)$ matrix whose columns form a basis for S and let Y be an $(n - r) \times n$ matrix whose rows form a basis for T. We claim that the bordered matrix

$$\begin{bmatrix} A & X \\ Y & 0 \end{bmatrix}$$

is nonsingular. Suppose

$$\begin{bmatrix} A & X \\ Y & 0 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Then $Au + Xv = 0$. Thus $Au = -Xv$ and in view of the choice of X, it follows that $Au = 0, Xv = 0$. Since the columns of X are linearly independent, $v = 0$. Also,

$Yu = 0$ and $Au = 0$ imply that $u = 0$. Hence the claim is proved. Let

$$\begin{bmatrix} A & X \\ Y & 0 \end{bmatrix}^{-1} = \begin{bmatrix} B & C \\ D & E \end{bmatrix}.$$

Then $AB + XD = I$, $AC + XE = 0$ and as before we conclude that $AC = 0$. We also have $BA + CY = I$. Thus $ABA + ACY = A$ and since $AC = 0$, we have $ABA = A$. Thus B is a g-inverse of A.

If we choose $S = R(A)^{\perp}$ and $T = R(A^T)^{\perp}$ then the g-inverse constructed above is the Moore–Penrose inverse.

Notes

The standard references for g-inverse are the classical books [2, 3 and 4]. The operator theoretic proof given in Proof 5 appears to be new. Further results on singular values and g-inverse, especially the Moore–Penrose inverse, are given in [1]. For g-inverses and bordered matrices, see [5].

Suggested Reading

[1] R B Bapat and Adi Ben-Israel, Singular values and maximum rank minors of generalized inverses, *Linear and Multilinear Algebra*, 40, 153-161, 1995.

[2] Adi Ben-Israel and T N E Greville, *Generalized Inverses: Theory and Applications*, Wiley-Interscience, New York, 1974.

[3] S L Campbell and C D Meyer Jr., *Generalized Inverses of Linear Transformations*, Pitman, 1979.

[4] Kentaro Nomakuchi, On characterization of generalized inverses, *Linear Algebra and Its Applications*, 33, 1-8, 1980.

[5] C R Rao and S K Mitra, *Generalized Inverse of Matrices and its Applications*, Wiley, New York, 1971.

Address for Correspondence
R B Bapat
Indian Statistical Institute
New Delhi 110 016, India

Existence of Generalized Inverse: Ten Proofs and Some Remarks

R B Bapat

R B Bapat is Professor at the Indian Statistical Institute, New Delhi. His main research areas have been nonnegative matrices, matrix inequalities, generalized inverses, and matrices in graph theory. He is author of *Linear algebra and linear models* (Hindustan Book Agency and Springer-Verlag) and, jointly with T E S Raghavan, *Nonnegative matrices and applications* (Cambridge University Press).

Introduction

The theory of g-inverses has seen a substantial growth over the past few decades. It is an area of great theoretical interest which finds applications in many diverse areas, including statistics, numerical analysis, Markov chains, differential equations and control theory.

What is the motivation for seeking a generalized notion of the inverse of a matrix? We briefly address this question now. Recall that an $n \times n$ matrix A is said to be nonsingular if there exists a matrix B such that $AB = I_n$, the identity matrix of order n. The inverse, when it exists, is unique and is denoted by A^{-1}. An $n \times n$ matrix A is nonsingular if and only if the associated linear transformation $f(x) = Ax$ is one-to-one. The concept of nonsingularity is of central importance in linear algebra and it easily relates to other concepts such as basis, rank, inner product, orthogonality and so on.

In practice, however, one often encounters matrices that are singular as well as matrices that are rectangular, rather than square, and hence clearly not nonsingular. The singularity seems to be inherent in the problem and cannot be made to go away by superficial means. As an example, in a statistical linear model one has a study variable Y which depends on some control variables $X_1, \ldots, X_p$. We set up the hypothesis that Y is a linear combination of $X_1, \ldots, X_p$ except for a random error. Often there is a linear relationship among $X_1, \ldots, X_p$ and this causes the covariance matrix of the errors corresponding to the observations on Y to be singular.