1

GENERALIZED INVERSE MATRICES

1. INTRODUCTION

Generalized inverse matrices are an important and useful mathematical tool for understanding certain aspects of the analysis procedures associated with linear models, especially the analysis of unbalanced data for non-full rank models. The analysis of unbalanced data and non-full rank models is of special importance and thus receives considerable attention in this book. Therefore, it is appropriate that we summarize the features of generalized inverses that are important to linear models. We will also discuss other useful and interesting results in matrix algebra.

We will frequently need to solve systems of equations of the form $\mathbf{Ax} = \mathbf{y}$ where $\mathbf{A}$ is an $m \times n$ matrix. When $m = n$ and $\mathbf{A}$ is nonsingular, the solution takes the form $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$.

For a consistent system of equations where m may not equal n, or for square singular matrices, there exist matrices $\mathbf{G}$ where $\mathbf{x} = \mathbf{Gy}$. These matrices are generalized inverses.

Example 1 Need for Generalized Inverses

Consider the system of equations

$$\begin{aligned} 5x_1 + 3x_2 + 2x_3 &= 50 \\ 3x_1 + 3x_2 &= 30 \\ 2x_1 + 2x_3 &= 20 \end{aligned}$$

Linear Models, Second Edition. Shayle R. Searle and Marvin H. J. Gruber.

or in matrix format

$$\begin{bmatrix} 5 & 3 & 2 \\ 3 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix}$$

Notice that the coefficient matrix is not of full rank. Indeed, the second and third rows add up to the first row. Solutions of this system include

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ 10 \end{bmatrix},$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{54}\begin{bmatrix} 5 & 1 & 4 \\ 1 & 11 & -10 \\ 4 & -10 & 14 \end{bmatrix}\begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix} = \begin{bmatrix} \frac{20}{3} \\ \frac{10}{3} \\ \frac{10}{3} \end{bmatrix}$$

and infinitely many others. Each of the 3×3 matrices in the above solutions is generalized inverses. □

a. Definition and Existence of a Generalized Inverse

In this book, we define a *generalized inverse* of a matrix $\mathbf{A}$ as any matrix $\mathbf{G}$ that satisfies the equation

$$\mathbf{AGA} = \mathbf{A}. \tag{1}$$

The reader may verify that the 3×3 matrices in the solutions to the system in Example 1 satisfies (1) and are thus, generalized inverses.

The name "generalized inverse" for matrices $\mathbf{G}$ defined by (1) is unfortunately not universally accepted. Names such as "conditional inverse," "pseudo inverse," and "g-inverse" are also to be found in the literature. Sometimes, these names refer to matrices defined as is $\mathbf{G}$ in (1) and sometimes to matrices defined as variants of $\mathbf{G}$. However, throughout this book, we use the name "generalized inverse" of $\mathbf{A}$ exclusively for any matrix $\mathbf{G}$ satisfying (1).

Notice that (1) does not define $\mathbf{G}$ as "the" generalized inverse of $\mathbf{A}$ but as "a" generalized inverse of $\mathbf{A}$. This is because **G,** for a given matrix, $\mathbf{A}$ is not unique. As we shall show below there is an infinite number of matrices $\mathbf{G}$ that satisfy (1). Thus, we refer to the whole class of them as generalized inverses of $\mathbf{A}$.

Notice that in Example 1, we gave two generalized inverses of the coefficient matrix of the system of equations. Lots more could have been found.

There are many ways to find generalized inverses. We will give three here.

The first starts with the equivalent diagonal form of **A**. If **A** has order $p \times q$, the reduction to this diagonal form can be written as

$$\mathbf{P}_{p\times p}\mathbf{A}_{p\times q}\mathbf{Q}_{q\times q} = \mathbf{\Delta}_{p\times q} \equiv \begin{bmatrix} \mathbf{D}_{r\times r} & \mathbf{0}_{r\times(q-r)} \\ \mathbf{0}_{(p-r)\times r} & \mathbf{0}_{(p-r)\times(q-r)} \end{bmatrix}$$

or more simply as

$$\mathbf{PAQ} = \mathbf{\Delta} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{2}$$

As usual, **P** and **Q** are products of elementary operators (see Searle, 1966, 2006, or Gruber, 2014). The matrix **A** has rank r and $\mathbf{D}_r$ is a diagonal matrix of order r. In general, if $d_1, d_2, \ldots, d_r$ are the diagonal elements of any diagonal matrix **D**, we will use the notation $\mathbf{D}\{d_i\}$for $\mathbf{D}_r$; that is,

$$\mathbf{D}_r \equiv \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ & & \ddots & \vdots \\ 0 & \cdots & 0 & d_r \end{bmatrix} \equiv \operatorname{diag}\{d_i\} = \mathbf{D}\{d_i\} \quad \text{for } i = 1, \ldots, r. \tag{3}$$

Furthermore, as in Δ, the symbol **0** will represent null matrices with order being determined by the context on each occasion.

Derivation of **G** comes easily from Δ. Analogous to Δ, we define Δ^- (to be read Δ minus) as

$$\mathbf{\Delta}^- = \begin{bmatrix} \mathbf{D}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Then as shown below

$$\mathbf{G} = \mathbf{Q\Delta^-P} \tag{4}$$

satisfies (1) and is thus a generalized inverse. The generalized inverse **G** as given by (4) is not unique, because neither **P** nor **Q** by their definition is unique, neither is Δ or Δ^-, and therefore $\mathbf{G} = \mathbf{Q}\Delta^-\mathbf{P}$ is not unique.

Before showing that **G** does satisfy (1), note from the definitions of Δ and Δ^- given above that

$$\mathbf{\Delta\Delta^-\Delta} = \mathbf{\Delta}. \tag{5}$$

Hence, by the definition implied in (1), we can say that Δ^- is a generalized inverse of Δ. While this is an unimportant result in itself, it enables us to establish that **G**, as defined in (3), is indeed a generalized inverse of $\mathbf{A}$. To show this, observe that from (2),

$$\mathbf{A} = \mathbf{P}^{-1}\boldsymbol{\Delta}\mathbf{Q}^{-1}. \tag{6}$$

The inverses of $\mathbf{P}$ and $\mathbf{Q}$ exist because $\mathbf{P}$ and $\mathbf{Q}$ are products of elementary matrices and are, as a result, nonsingular. Then from (4), (5), and (6), we have,

$$\mathbf{AGA} = \mathbf{P}^{-1}\boldsymbol{\Delta}\mathbf{Q}^{-1}\mathbf{Q}\boldsymbol{\Delta}^{-}\mathbf{P}\mathbf{P}^{-1}\boldsymbol{\Delta}\mathbf{Q}^{-1} = \mathbf{P}^{-1}\boldsymbol{\Delta}\boldsymbol{\Delta}^{-}\boldsymbol{\Delta}\mathbf{Q}^{-1} = \mathbf{P}^{-1}\boldsymbol{\Delta}\mathbf{Q}^{-1} = \mathbf{A}. \tag{7}$$

Thus, (1) is satisfied and $\mathbf{G}$ is a generalized inverse of $\mathbf{A}$.

Example 2 Obtaining a Generalized Inverse by Matrix Diagonalization
For

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 1 & 5 \\ 3 & 1 & 3 \end{bmatrix},$$

a diagonal form is obtained using

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 0 \\ -\frac{2}{3} & -\frac{1}{3} & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Q} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus,

$$\mathbf{PAQ} = \boldsymbol{\Delta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Delta}^{-} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

As a result,

$$\mathbf{G} = \mathbf{Q}\boldsymbol{\Delta}^{-}\mathbf{P} = \frac{1}{3}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The reader may verify that $\mathbf{AGA} = \mathbf{A}$. □

It is to be emphasized that generalized inverses exist for rectangular matrices as well as for square ones. This is evident from the formulation of $\boldsymbol{\Delta}_{p\times q}$. However, for $\mathbf{A}$ of order $p \times q$, we define $\boldsymbol{\Delta}^{-}$ as having order $q \times p$, the null matrices therein being of appropriate order to make this so. As a result, the generalized inverse **G** has order $q \times p$.

Example 3 Generalized Inverse for a Matrix That Is Not Square
Consider

$$\mathbf{B} = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

the same **A** in the previous example with an additional column With **P** as given in Example 2 and **Q** now taken as

$$\mathbf{Q} = \begin{bmatrix} 1 & -1 & 1 & 5 \\ 0 & 1 & -6 & -20 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{PBQ} = \boldsymbol{\Delta} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We then have

$$\boldsymbol{\Delta}^- = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ so that } \mathbf{G} = \mathbf{Q}\boldsymbol{\Delta}^-\mathbf{P} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{4}{3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

□

b. An Algorithm for Obtaining a Generalized Inverse

The algorithm is based on knowing or first finding the rank of the matrix. We present the algorithm first and then give a rationale for why it works. The algorithm goes as follows:

1. In **A** of rank r, find any non-singular minor of order r. Call it **M**.
2. Invert **M** and transpose the inverse to obtain $(\mathbf{M}^{-1})'$.
3. In **A,** replace each element of **M** by the corresponding element of $(\mathbf{M}^{-1})'$.
4. Replace all other elements of **A** by zero.
5. Transpose the resulting matrix.

The result is a generalized inverse of **A**. Observe that different choices of the minor of rank r will give different generalized inverses of **A**.

Example 4 Computing a Generalized Inverse using the Algorithm

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 & 2 \\ 3 & 7 & 12 & 4 \\ 0 & 1 & -3 & -2 \end{bmatrix}.$$

The reader may verify that all of the 3×3 sub-matrices of $\mathbf{A}$ have determinant zero while the 2×2 sub-matrices have non-zero determinants. Thus, $\mathbf{A}$ has rank 2. Consider

$$\mathbf{M} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}.$$

Then

$$\mathbf{M}^{-1} = \begin{bmatrix} 7 & -2 \\ -3 & 1 \end{bmatrix}$$

and

$$(\mathbf{M}^{-1})' = \begin{bmatrix} 7 & -3 \\ -2 & 1 \end{bmatrix}.$$

Now write the matrix

$$\mathbf{H} = \begin{bmatrix} 7 & -3 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then the generalized inverse

$$\mathbf{G} = \mathbf{H}' = \begin{bmatrix} 7 & -2 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

By a similar process, taking

$$\mathbf{M} = \begin{bmatrix} 12 & 4 \\ -3 & -2 \end{bmatrix},$$

another generalized inverse of **A** is

$$\tilde{\mathbf{G}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{3} \\ 0 & -\frac{1}{4} & -1 \end{bmatrix}.$$

The reader may, if he/she wishes, construct other generalized inverses using 2×2 sub-matrices with non-zero determinant. □

We now present the rationale for the algorithm. Suppose **A** can be partitioned in such a way that its leading $r \times r$ minor is non-singular, that is,

$$\mathbf{A}_{p\times q} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}$ is $r \times r$ of rank r. Then a generalized inverse of **A** is

$$\mathbf{G}_{q\times p} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

where the null matrices are of appropriate order to make **G** a $q \times p$ matrix. To see that **G** is a generalized inverse of **A**, note that

$$\mathbf{AGA} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}.$$

Now since **A** is of rank r, the rows are linearly dependent. Thus, for some matrix $\mathbf{K}\ [\mathbf{A}_{21}\quad \mathbf{A}_{22}] = \mathbf{K}[\mathbf{A}_{11}\quad \mathbf{A}_{12}]$. Specifically $\mathbf{K} = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}$ and so $\mathbf{A}_{22} = \mathbf{KA}_{12} = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. Hence, $\mathbf{AGA} = \mathbf{A}$ and **G** is a generalized inverse of **A**.

There is no need for the non-singular minor to be in the leading position. Let **R** and **S** represent the elementary row and column operations, respectively, to bring it to the leading position. Then **R** and **S** are products of elementary operators with

$$\mathbf{RAS} = \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \tag{8}$$

where $\mathbf{B}_{11}$ is non-singular of order r. Then

$$\mathbf{F} = \begin{bmatrix} \mathbf{B}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

is a generalized inverse of $\mathbf{B}$ and $\mathbf{G}_{q\times p} = \mathbf{SFR}$ is a generalized inverse of A. From (8),

$$\mathbf{A} = \mathbf{R}^{-1}\mathbf{B}\mathbf{S}^{-1}.$$

Then

$$\mathbf{AGA} = \mathbf{R}^{-1}\mathbf{B}\mathbf{S}^{-1}\mathbf{SFR}\mathbf{R}^{-1}\mathbf{B}\mathbf{S}^{-1} = \mathbf{R}^{-1}\mathbf{BFB}\mathbf{S}^{-1} = \mathbf{R}^{-1}\mathbf{B}\mathbf{S}^{-1} = \mathbf{A}.$$

Now $\mathbf{R}$ and $\mathbf{S}$ are products of elementary operators that exchange rows and columns. Such matrices are identity matrices with rows and columns interchanged. Such matrices are known as *permutation matrices* and are orthogonal. Thus, we have that $\mathbf{R} = \mathbf{I}$ with its rows in a different sequence, a permutation matrix and $\mathbf{R}'\mathbf{R} = \mathbf{I}$. The same is true for S and so from (8), we have that

$$\mathbf{A} = \mathbf{R}'\mathbf{B}\mathbf{S}' = \mathbf{R}'\begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}\mathbf{S}'. \tag{9}$$

As far as $\mathbf{B}_{11}$ is concerned, the product in (9) represents the operations of returning the elements of $\mathbf{B}_{11}$ to their original position in $\mathbf{A}$. Now consider $\mathbf{G}$. We have

$$\mathbf{G} = \mathbf{SFR} = (\mathbf{R}'\mathbf{F}'\mathbf{S}')' = \left\{ \mathbf{R}' \begin{bmatrix} (\mathbf{B}_{11}^{-1})' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{S}' \right\}$$

In this, analogous to the form of $\mathbf{A} = \mathbf{R}'\mathbf{B}\mathbf{S}'$ the product involving $\mathbf{R}'$ and $\mathbf{S}'$ in $\mathbf{G}'$ represents putting the elements of $(\mathbf{B}_{11}^{-1})'$ into the corresponding positions of $\mathbf{G}'$ that the elements of $\mathbf{B}_{11}$ occupied in $\mathbf{A}$. This is what motivates the algorithm.

c. Obtaining Generalized Inverses Using the Singular Value Decomposition (SVD)

Let $\mathbf{A}$ be a matrix of rank r. Let $\mathbf{\Lambda}$ be $r \times r$ the diagonal matrix of non-zero eigenvalues of $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ ordered from largest to smallest. The non-zero eigenvalues of $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ are the same (see p. 110 of Gruber (2014) for a proof). Then the decomposition of

$$\mathbf{A} = \begin{bmatrix} \mathbf{S}' & \mathbf{T}' \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}' \\ \mathbf{V}' \end{bmatrix} = \mathbf{S}'\mathbf{\Lambda}^{1/2}\mathbf{U}', \tag{10}$$

where $[\,\mathbf{S}' \quad \mathbf{T}'\,]$ and $[\,\mathbf{U} \quad \mathbf{V}\,]$ are orthogonal matrices, is the singular value decomposition (SVD). The existence of this decomposition is established in Gruber (2014) following Stewart (1963, p. 126). Observe that $\mathbf{S}'\mathbf{S} + \mathbf{T}'\mathbf{T} = \mathbf{I}, \mathbf{UU}' + \mathbf{VV}' = \mathbf{I}, \mathbf{SS}' = \mathbf{I}, \mathbf{TT}' = \mathbf{I}, \mathbf{S}'\mathbf{T} = \mathbf{0}, \mathbf{T}'\mathbf{S} = \mathbf{0}, \mathbf{UU}' = \mathbf{I}, \mathbf{U}'\mathbf{V} = \mathbf{0},$ and $\mathbf{V}'\mathbf{U} = \mathbf{0}$. Furthermore, $\mathbf{A}'\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$ and $\mathbf{A}\mathbf{A}' = \mathbf{S}'\mathbf{\Lambda}\mathbf{S}$. A generalized inverse of $\mathbf{A}$ then takes the form

$$\mathbf{G} = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{S}. \tag{11}$$

Indeed, $\mathbf{AGA} = \mathbf{S}'\mathbf{\Lambda}^{1/2}\mathbf{U}'\mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{SS}'\mathbf{\Lambda}^{1/2}\mathbf{U}' = \mathbf{S}'\mathbf{\Lambda}^{1/2}\mathbf{U}' = \mathbf{A}$.

Example 5 Finding a Generalized Inverse using the Singular Value Decomposition Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Then,

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{A}\mathbf{A}' = \begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}.$$

To find the eigenvalues of $\mathbf{A}'\mathbf{A}$ solve the equation

$$\det \begin{bmatrix} 4-\lambda & 2 & 2 \\ 2 & 2-\lambda & 0 \\ 2 & 0 & 2-\lambda \end{bmatrix} = 0$$

or

$$\lambda^3 - 8\lambda^2 + 12\lambda = \lambda(\lambda - 6)(\lambda - 2) = 0$$

to get the eigenvalues $\lambda = 6, 2, 0$. Finding the eigenvectors by solving the systems of equations

$$\begin{array}{c} -2x_1 + 2x_2 + 2x_3 = 0 \\ 2x_1 - 4x_2 = 0 \\ 2x_1 - 4x_3 = 0 \end{array} \qquad \begin{array}{c} 2x_1 + 2x_2 + 2x_3 = 0 \\ 2x_1 = 0 \end{array} \qquad \begin{array}{c} 4x_1 + 2x_2 + 2x_3 = 0 \\ 2x_1 + 2x_2 = 0 \\ 2x_1 + 2x_3 = 0 \end{array}$$

yields a matrix of normalized eigenvectors of $\mathbf{A}'\mathbf{A}$,

$$[\,\mathbf{U} \quad \mathbf{V}\,] = \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

By a similar process, the reader may show that the eigenvalues of $\mathbf{AA}'$ are $\lambda = 6, 2, 0, 0$ and that the matrix of eigenvectors is

$$\begin{bmatrix}\mathbf{S}' & \mathbf{T}'\end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

Then the singular value decomposition of

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and, as a result, the generalized inverse

$$\mathbf{G} = \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

□

These derivations of a generalized inverse matrix **G** are by no means the only ways such a matrix can be computed. For matrices of small order, they can be satisfactory, but for those of large order that might occur in the analysis of "big data," other methods might be preferred. Some of these are discussed subsequently. Most methods involve, of course, the same kind of numerical problems as are incurred in calculating the regular inverse $\mathbf{A}^{-1}$ of a non-singular matrix **A**. Despite this, the generalized inverse has importance because of its general application to non-square matrices and to square singular matrices. In the special case that **A** is non-singular, $\mathbf{G} = \mathbf{A}^{-1}$ as one would expect, and in this case, **G** is unique.

The fact that $\mathbf{A}$ has a generalized inverse even when it is singular or rectangular has particular application in the problem of solving equations, for example, of solving $\mathbf{Ax} = \mathbf{y}$ for $\mathbf{x}$ when $\mathbf{A}$ is singular or rectangular. In situations of this nature, the use of a generalized inverse $\mathbf{G}$, as we shall see, leads very directly to a solution in the form $\mathbf{x} = \mathbf{Gy}$. This is of great importance in the study of linear models where such equations arise quite frequently. For example, when we can write a linear model as $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$, finding the least square estimator for estimating $\mathbf{b}$ leads to equations $\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}$ where the matrix $\mathbf{X}'\mathbf{X}$ is singular. Hence, we cannot write the solution as $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. However, using a generalized inverse $\mathbf{G}$ of $\mathbf{X}'\mathbf{X}$, we can obtain the solution directly in the form $\mathbf{GX}'\mathbf{y}$ and study its properties.

For linear models, the use of generalized inverse matrices in solving linear equations is the application of prime interest. We now outline the resulting procedures. Following this, we discuss some general properties of generalized inverses.

2. SOLVING LINEAR EQUATIONS

a. Consistent Equations

A convenient starting point from which to develop the solution of linear equations using a generalized inverse is the definition of consistent equations.

Definition 1 The linear equations $\mathbf{Ax} = \mathbf{y}$ are defined as being consistent if any linear relationships existing among the rows of $\mathbf{A}$ also exist among the corresponding elements of $\mathbf{y}$. In other words, $\mathbf{t}'\mathbf{A} = 0$ if and only if $\mathbf{t}'\mathbf{y} = 0$ for any vector $\mathbf{t}$.

As a simple example, the equations

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 21 \end{bmatrix}$$

are consistent. The second row of the matrix on the left-hand side of the system is the first row multiplied and on the right-hand side, of course $21 = 7(3)$. On the other hand, the equations

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 24 \end{bmatrix}$$

are inconsistent. The linear relationship between the rows of the matrix on the left-hand side of the system does not hold true between 7 and 24. Moreover, you can write out the two equations and show that $3 = 0$.

The formal definition of consistent equations does not demand that linear relationships exist among the rows of $\mathbf{A}$. However, if they do, then the definition does require that the same relationships also exist among the corresponding elements of $\mathbf{y}$ for the equations to be consistent. For example, when $\mathbf{A}$ is non-singular, the equations

$\mathbf{Ax} = \mathbf{y}$ are always consistent. There are no linear relationships among the rows of $\mathbf{A}$ and therefore none that the elements of $\mathbf{y}$ must satisfy.

The importance of consistency lies in the following theorem. Linear equations can be solved only if they are consistent. See, for example, Section 6.2 of Searle (1966) or Section 7.2 of Searle and Hausman (1970) for a proof. Since only consistent equations can be solved, discussion of a procedure for solving linear equations is hereafter confined to equations that are consistent. The procedure is described in Theorems 1 and 2 in Section 2b. Theorems 3–6 in Section 2c deal with the properties of these solutions.

b. Obtaining Solutions

The link between a generalized inverse of the matrix $\mathbf{A}$ and consistent equations $\mathbf{Ax} = \mathbf{y}$ is set out in the following theorem adapted from C. R. Rao (1962).

Theorem 1 Consistent equations $\mathbf{Ax} = \mathbf{y}$ have a solution $\mathbf{x} = \mathbf{Gy}$ if and only if $\mathbf{AGA} = \mathbf{A}$.

Proof. If the equations $\mathbf{Ax} = \mathbf{y}$ are consistent and have $\mathbf{x} = \mathbf{Gy}$ as a solution, write $\mathbf{a}_j$ for the jth column of $\mathbf{A}$ and consider the equations $\mathbf{Ax} = \mathbf{a}_j$. They have a solution. It is the null vector with its jth element set equal to unity. Therefore, the equations $\mathbf{Ax} = \mathbf{a}_j$ are consistent. Furthermore, since consistent equations $\mathbf{Ax} = \mathbf{y}$ have a solution $\mathbf{x} = \mathbf{Gy}$, it follows that consistent equations $\mathbf{Ax} = \mathbf{a}_j$ have a solution $\mathbf{x} = \mathbf{Ga}_j$. Therefore,

$\mathbf{AGa}_j = \mathbf{a}_j$ and this is true for all values of j, that is, for all columns of $\mathbf{A}$. Hence, $\mathbf{AGA} = \mathbf{A}$.

Conversely, if $\mathbf{AGA} = \mathbf{A}$ then $\mathbf{AGAx} = \mathbf{Ax}$, and when $\mathbf{Ax} = \mathbf{y}$ substitution gives $\mathbf{A}(\mathbf{Gy}) = \mathbf{y}$. Hence, $\mathbf{x} = \mathbf{Gy}$ is a solution of $\mathbf{Ax} = \mathbf{y}$ and the theorem is proved. ■

Theorem 1 indicates how a solution to consistent equations may be obtained. Find any generalized inverse of $\mathbf{A}$, $\mathbf{G}$, and then $\mathbf{Gy}$ is a solution. However, this solution is not unique. There are, indeed, many solutions whenever $\mathbf{A}$ is anything but a square, non-singular matrix. These are characterized in Theorem 2 and 3.

Theorem 2 If $\mathbf{A}$ has q columns and $\mathbf{G}$ is a generalized inverse of $\mathbf{A}$, then the consistent equations $\mathbf{Ax} = \mathbf{y}$ have the solution

$$\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z} \tag{12}$$

where $\mathbf{z}$ is any arbitrary vector of order q.

Proof. Since $\mathbf{AGA} = \mathbf{A}$, $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{AGy} + (\mathbf{AGA} - \mathbf{A})\mathbf{z} = \mathbf{AGy} = \mathbf{y}$, by Theorem 1. ■

There are as many solutions to (12) as there are choices of $\mathbf{z}$ and $\mathbf{G}$. Thus, the equation $\mathbf{Ax} = \mathbf{y}$ has infinitely many solutions of the form (12).

Example 6 Different Solutions to Ax = y for a particular A

Consider the equations $\mathbf{Ax} = \mathbf{y}$ **as**

$$\begin{bmatrix} 5 & 3 & 1 & -4 \\ 8 & 5 & 2 & 3 \\ 21 & 13 & 5 & 2 \\ 3 & 2 & 1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 22 \\ 2 \end{bmatrix}, \tag{13}$$

so defining $\mathbf{A}$, $\mathbf{x}$, and $\mathbf{y}$. Using the algorithm developed in Section 1b with the 2×2 minor in the upper left-hand corner of $\mathbf{A}$, it will be found that

$$\mathbf{G} = \begin{bmatrix} 5 & -3 & 0 & 0 \\ -8 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$. The solution of the form (12) is

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z} \\ &= \begin{bmatrix} 6 \\ -8 \\ 0 \\ 0 \end{bmatrix} + \left\{ \begin{bmatrix} 1 & 0 & -1 & -29 \\ 0 & 1 & 2 & 47 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \mathbf{I} \right\} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} \\ &= \begin{bmatrix} 6 - z_3 - 29z_4 \\ -8 + 2z_3 + 47z_4 \\ -z_3 \\ -z_4 \end{bmatrix} \end{aligned} \tag{14}$$

where z_3 and z_4 are arbitrary. This means that (13) is a solution to (12) no matter what the given values of z_3 and z_4 are. For example putting $z_3 = z_4 = 0$ gives

$$\tilde{\mathbf{x}}_1' = \begin{bmatrix} 6 & -8 & 0 & 0 \end{bmatrix} \tag{15}$$

Setting $z_3 = -1$ and $z_4 = 2$ gives

$$\tilde{\mathbf{x}}_2' = \begin{bmatrix} -51 & 84 & 1 & -2 \end{bmatrix}. \tag{16}$$

Both of the results in (15) and (16) can be shown to satisfy (13) by direct substitution.

This is also true of the result in (14) for all z_3 and z_4.

Again, using the algorithm in Section 1b, this time using the 2×2 minor in the second and third row and column, we obtain the generalized inverse

$$\dot{\mathbf{G}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -5 & 2 & 0 \\ 0 & 13 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then (12) becomes

$$\begin{aligned} \mathbf{x} &= \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z} \\ &= \begin{bmatrix} 0 \\ 4 \\ -6 \\ 0 \end{bmatrix} + \left\{ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & -11 \\ -1 & 0 & 1 & 29 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \mathbf{I} \right\} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} \\ &= \begin{bmatrix} -\dot{z}_1 \\ 4 + 2\dot{z}_1 - 11\dot{z}_4 \\ -6 - \dot{z}_1 + 29\dot{z}_4 \\ -\dot{z}_4 \end{bmatrix} \end{aligned} \tag{17}$$

for arbitrary values $\dot{z}_1$ and $\dot{z}_4$. The reader may show that this too satisfies (13). □

c. Properties of Solutions

One might ask about the relationship, if any, between the two solutions (14) and (17) found by using the two generalized inverses $\mathbf{G}$ and $\dot{\mathbf{G}}$. Both satisfy (13) for an infinite number of sets of values of z_3, z_4 and $\dot{z}_1, \dot{z}_4$. The basic question is: do the two solutions generate, though allocating different sets of values to the arbitrary values z_3 and z_4 in $\tilde{\mathbf{x}}$ and $\dot{z}_1$ and $\dot{z}_4$ in $\dot{\mathbf{x}}$, the same series of vectors satisfying $\mathbf{Ax} = \mathbf{y}$? The answer is "yes" because on substituting $\dot{z}_1 = -6 + z_3 + 29z_4$ and $\dot{z}_4 = z_4$ into (17) yields the solution in (14). Hence, (14) and (17) generate the same sets of solutions.

Likewise, the relationship between solutions using $\mathbf{G}$ and those using $\dot{\mathbf{G}}$ is that on substituting $\mathbf{z} = (\mathbf{G} - \dot{\mathbf{G}})\mathbf{y} + (\mathbf{I} - \dot{\mathbf{G}}\mathbf{A})\dot{\mathbf{z}}$ into (12) and noting by Theorem 1 that $\mathbf{GAGy} = \mathbf{GA}\dot{\mathbf{G}}\mathbf{y}$ $\tilde{\mathbf{x}}$ reduces to $\dot{\mathbf{x}}$.

A stronger result which concerns generation of all solutions from $\tilde{\mathbf{x}}$ is contained in the following theorem.

Theorem 3 For the consistent equations $\mathbf{Ax} = \mathbf{y}$, all solutions are, for any specific $\mathbf{G}$ generated by $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z}$ for arbitrary $\mathbf{z}$.

Proof. Let $\mathbf{x}^*$ be any solution to $\mathbf{Ax} = \mathbf{y}$. Choose $\mathbf{z} = (\mathbf{GA} - \mathbf{I})\mathbf{x}^*$. Then

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})(\mathbf{GA} - \mathbf{I})\mathbf{x}^* \\ &= \mathbf{Gy} + (\mathbf{GAGA} - \mathbf{GA} - \mathbf{GA} + \mathbf{I})\mathbf{x}^* \\ &= \mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{x}^* = \mathbf{Gy} + \mathbf{x}^* - \mathbf{GAx}^* \\ &= \mathbf{Gy} + \mathbf{x}^* - \mathbf{Gy} = \mathbf{x}^* \end{aligned}$$

applying Theorem 1. ■

The importance of this theorem is that we need to derive only one generalized inverse of $\mathbf{A}$ to be able to generate all solutions to $\mathbf{Ax} = \mathbf{y}$. There are no solutions other than those that can be generated from $\tilde{\mathbf{x}}$.

Having established a method for solving linear equations and showing that they can have an infinite number of solutions, we ask two questions:

(i) What relationships exist among the solutions?
(ii) To what extent are the solutions linearly independent (LIN)? (A discussion of linear independence and dependence is available in Section 5 of Gruber (2014) or any standard matrix or linear algebra textbook.)

Since each solution is a vector of order q, there can of course be no more than q LIN solutions. In fact, there are fewer, as Theorem 4 shows.

Theorem 4 When $\mathbf{A}$ is a matrix of q columns and rank r, and when $\mathbf{y}$ is a non-null vector, the number of LIN solutions to the consistent equations $\mathbf{Ax} = \mathbf{y}$ is $q - r + 1$.

To establish this theorem we need the following Lemma.

Lemma 1 Let $\mathbf{H} = \mathbf{GA}$ where the rank of $\mathbf{A}$, denoted by $r(\mathbf{A})$ is r, that is, $r(\mathbf{A}) = r$; and $\mathbf{A}$ has q columns. Then $\mathbf{H}$ is idempotent (meaning that $\mathbf{H}^2 = \mathbf{H}$) with rank r and

$$r(\mathbf{I} - \mathbf{H}) = q - r.$$

Proof. To show that $\mathbf{H}$ is idempotent, notice that $\mathbf{H}^2 = \mathbf{GAGA} = \mathbf{GA} = \mathbf{H}$. Furthermore, by the rule for the rank of a product matrix (See Section 6 of Gruber (2014)), $r(\mathbf{H}) = r(\mathbf{GA}) \leq r(\mathbf{A})$. Similarly, because $\mathbf{AH} = \mathbf{AGA} = \mathbf{A}$, $r(\mathbf{H}) \geq r(\mathbf{AH}) = r(\mathbf{A})$. Therefore, $r(\mathbf{H}) = r(\mathbf{A}) = r$. Since $\mathbf{H}$ is idempotent, we have that $(\mathbf{I} - \mathbf{H})^2 = \mathbf{I} - 2\mathbf{H} + \mathbf{H}^2 = \mathbf{I} - 2\mathbf{H} + \mathbf{H} = \mathbf{I} - \mathbf{H}$. Thus, $\mathbf{I} - \mathbf{H}$ is also idempotent of order q. The eigenvalues of an idempotent matrix can be shown to be zero or one. The rank of a matrix corresponds to the number of non-zero eigenvalues. The trace of an idempotent matrix is the number of non-zero eigenvalues. Thus, $r(\mathbf{I} - \mathbf{H}) = \mathrm{tr}(\mathbf{I} - \mathbf{H}) = q - \mathrm{tr}(\mathbf{H}) = q - r$. ■

Proof of Theorem 4. Writing $\mathbf{H} = \mathbf{GA}$, the solutions to $\mathbf{Ax} = \mathbf{y}$ are from Theorem 2

$$\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z}.$$

From Lemma 1, $r(\mathbf{I} - \mathbf{H}) = q - r$. As a result, there are only $(q - r)$ arbitrary elements in $(\mathbf{H} - \mathbf{I})\mathbf{z}$. The other r elements are linear combinations of those $q - r$. Therefore, there only $(q - r)$ LIN vectors $(\mathbf{H} - \mathbf{I})\mathbf{z}$ and using them in $\tilde{\mathbf{x}}$ gives $(q - r)$ LIN solutions. For $i = 1, 2, \ldots, q - r$ let $\tilde{\mathbf{x}}_i = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i$ be these solutions. Another solution is $\tilde{\mathbf{x}} = \mathbf{Gy}$.

Assume that this solution is linearly dependent on the $\tilde{\mathbf{x}}_i$. Then, for scalars $\lambda_i,\ i = 1,\ 2, \dots ,\ q - r$, not all of which are zero,

$$\begin{aligned}\mathbf{Gy} &= \sum_{i=1}^{q-r} \lambda_i \tilde{\mathbf{x}}_i = \sum_{i=1}^{q-r} \lambda_i [\mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i] \\ &= \mathbf{Gy} \sum_{i=1}^{q-r} \lambda_i + \sum_{i=1}^{q-r} \lambda_i [(\mathbf{H} - \mathbf{I})\mathbf{z}_i].\end{aligned} \tag{18}$$

The left-hand side of (18) contains no $\mathbf{z}$'s. Therefore, for the last expression on the right-hand side of (18), the second term is zero. However, since the $(\mathbf{H} - \mathbf{I})\mathbf{z}_i$ are LIN, this can be true only if each of the λ_i is zero. This means that (18) is no longer true for some λ_i non-zero. Therefore, $\mathbf{Gy}$ is independent of the $\tilde{\mathbf{x}}_i$ so that $\mathbf{Gy}$ and $\tilde{\mathbf{x}}_i$ for $i = 1,\ 2, \dots ,\ q - r$ form a set of $(q - r + 1)$ LIN solutions. When $q = r$, there is but one solution corresponding to the existence of $\mathbf{A}^{-1}$, and that one solution is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. ■

Theorem 4 means that $\tilde{\mathbf{x}} = \mathbf{Gy}$ and $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{H} - \mathbf{I})\mathbf{z}$ for $(q - r)$ LIN vectors $\mathbf{z}$ are LIN solutions to $\mathbf{Ax} = \mathbf{y}$. All other solutions will be linear combinations of these $(q - r + 1)$ solutions. Theorem 5 presents a way of constructing solutions in terms of other solutions.

Theorem 5 If $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots , \tilde{\mathbf{x}}_s$ are any s solutions of the consistent equations $\mathbf{Ax} = \mathbf{y}$ for which $\mathbf{y} \neq \mathbf{0}$, then any linear combination of these equations $\mathbf{x}^* = \sum_{i=1}^{s} \lambda_i \tilde{\mathbf{x}}_i$ is also a solution of the equations if and only if $\sum_{i=1}^{s} \lambda_i = 1$.

Proof. Since

$$\mathbf{x}^* = \sum\nolimits_{i=1}^{s} \lambda_i \tilde{\mathbf{x}}_i,$$

it follows that

$$\mathbf{Ax}^* = \mathbf{A} \sum_{i=1}^{s} \lambda_i \tilde{\mathbf{x}}_i = \sum_{i=1}^{s} \lambda_i \mathbf{A}\tilde{\mathbf{x}}_i.$$

Since $\tilde{\mathbf{x}}_i$ is a solution, for all i, $\mathbf{A}\tilde{\mathbf{x}}_i = \mathbf{y}$. This yields

$$\mathbf{Ax}^* = \sum_{i=1}^{s} \lambda_i \mathbf{y} = \mathbf{y} \left(\sum_{i=1}^{s} \lambda_i \right). \tag{19}$$

Now if $\mathbf{x}^*$ is a solution of $\mathbf{Ax} = \mathbf{y}$, then $\mathbf{Ax}^* = \mathbf{y}$ and by comparison with (19), this means, $\mathbf{y}$ being non-null, that $\sum_{i=1}^{s} \lambda_i = 1$. Conversely, if $\sum_{i=1}^{s} \lambda_i = 1$, equation (19) implies that $\mathbf{Ax}^* = \mathbf{y}$, meaning that $\mathbf{x}^*$ is a solution. This establishes the theorem. ■

Notice that Theorem 5 is in terms of any s solutions. Hence, for any number of solutions whether LIN or not, any linear combination of them is itself a solution provided that the coefficients in that combination sum to unity.

Corollary 5.1 When $\mathbf{y} = \mathbf{0}, \mathbf{Gy} = 0$ and there are only $q - r$ LIN solutions to $\mathbf{Ax} = \mathbf{0}.$ Furthermore, for any values of the λ_i's, $\sum_{i=1}^{s} \lambda_i \tilde{\mathbf{x}}_i$ is a solution to $\mathbf{Ax} = \mathbf{0}.$

Example 7 Continuation of Example 6

For $\mathbf{A}$ defined in Example 6, the rank $r = 2$. Therefore, there are $q - r + 1 = 4 - 2 + 1 = 3$ LIN solutions to (13). Two are shown in (14) and (15) with (14) being the solution $\mathbf{Gy}$ when the value $\mathbf{z} = \mathbf{0}$ is used. Another solution putting $\mathbf{z}' = \begin{bmatrix} 0 & 0 & -1 & 0 \end{bmatrix}$ into (14) is

$$\tilde{\mathbf{x}}_3' = \begin{bmatrix} -23 & 39 & 0 & -1 \end{bmatrix}.$$

Thus, $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2,$ and $\tilde{\mathbf{x}}_3$ are LIN solutions and any other solution will be a combination of these three. For example, with $\mathbf{z}' = \begin{bmatrix} -23 & 39 & 0 & -1 \end{bmatrix}$, the solution (14) becomes

$$\tilde{\mathbf{x}}_4 = \begin{bmatrix} 7 & -10 & 1 & 0 \end{bmatrix}.$$

It can be seen that

$$\tilde{\mathbf{x}}_4 = 2\tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2 - 2\tilde{\mathbf{x}}_3.$$

The coefficients on the right-hand side of the above linear combination sums to unity in accordance with Theorem 5.

A final theorem is related to an invariance property of the elements of a solution. It is important to the study of linear models because of its relationship with the concept of estimability discussed in Chapter 5. Without worrying about the details of estimability here, we give the theorem and refer to it later as needed. The theorem is due to C. R. Rao (1962). It concerns linear combinations of the elements of a solution vector. Certain combinations are invariant to whatever solution is used.

Theorem 6 The value of $\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to whatever solution is of $\mathbf{Ax} = \mathbf{y}$ is used for $\tilde{\mathbf{x}}$ if and only if $\mathbf{k}'\mathbf{H} = \mathbf{k}'$(where $\mathbf{H} = \mathbf{GA}$ and $\mathbf{AGA} = \mathbf{A}$).

Proof. For a solution $\tilde{\mathbf{x}}$ given by Theorem 2

$$\mathbf{k}'\tilde{\mathbf{x}} = \mathbf{k}'\mathbf{Gy} + \mathbf{k}'(\mathbf{H} - \mathbf{I})\mathbf{z}.$$

This is independent of the arbitrary $\mathbf{z}$ if $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. Since any solution can be put in the form $\tilde{\mathbf{x}}$ by the appropriate choice of $\mathbf{z}$, the value of $\mathbf{k}'\tilde{\mathbf{x}}$ for any $\tilde{\mathbf{x}}$ is $\mathbf{k}'\mathbf{G}\mathbf{y}$ provided that $\mathbf{k}'\mathbf{H} = \mathbf{k}'$.

It may not be entirely clear that when $\mathbf{k}'\mathbf{H} = \mathbf{k}'$, the value of $\mathbf{k}'\tilde{\mathbf{x}} = \mathbf{k}'\mathbf{G}\mathbf{y}$ is invariant to the choice of $\mathbf{G}$. We therefore clarify this point. First, by Theorem 4, there are $(q - r + 1)$ LIN solutions of the form $\tilde{\mathbf{x}} = \mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}$. Let these solutions be $\tilde{\mathbf{x}}_i$ for $i = 1,\ 2, \dots,\ q - r + 1$. Suppose that for some other generalized inverse, $\mathbf{G}^*$ we have a solution

$$\mathbf{x}^* = \mathbf{G}^*\mathbf{y} + (\mathbf{H}^* - \mathbf{I})\mathbf{z}^*.$$

Then, since the $\tilde{\mathbf{x}}_i$ are a LIN set of $(q - r + 1)$ solutions $\mathbf{x}^*$ must be a linear combination of them. This means that there is a set of scalars λ_i for $i = 1, 2, \dots, q - r + 1$ such that

$$\mathbf{x}^* = \sum_{i=1}^{q-r+1} \lambda_i \tilde{\mathbf{x}}_i$$

where not all of the λ_i s are zero. Furthermore, by Theorem 5, $\sum_{i=1}^{q-r+1} \lambda_i = 1$.

Proving the sufficiency part of the theorem demands showing that $\mathbf{k}'\tilde{\mathbf{x}}$ is the same for all solutions $\tilde{\mathbf{x}}$ when $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. Note that when $\mathbf{k}'\mathbf{H} = \mathbf{k}'$,

$$\mathbf{k}'\tilde{\mathbf{x}} = \mathbf{k}'\mathbf{H}\tilde{\mathbf{x}} = \mathbf{k}'\mathbf{H}\mathbf{G}\mathbf{y} + \mathbf{k}'(\mathbf{H}^2 - \mathbf{H})\mathbf{z} = \mathbf{k}'\mathbf{H}\mathbf{G}\mathbf{y} = \mathbf{k}'\mathbf{G}\mathbf{y}.$$

Therefore, $\mathbf{k}'\tilde{\mathbf{x}}_i = \mathbf{k}'\mathbf{G}\mathbf{y}$ for all i, and

$$\begin{aligned}\mathbf{k}'\mathbf{x}^* &= \mathbf{k}'\sum_{i=1}^{q-r+1} \lambda_i \tilde{\mathbf{x}}_i = \sum_{i=1}^{q-r+1} \lambda_i \mathbf{k}\tilde{\mathbf{x}}_i = \sum_{i=1}^{q-r+1} \lambda_i \mathbf{k}\mathbf{G}\mathbf{y} = \mathbf{k}'\mathbf{G}\mathbf{y}\left(\sum_{i=1}^{q-r+1} \lambda_i\right) \\ &= \mathbf{k}'\mathbf{G}\mathbf{y} = \mathbf{k}'\tilde{\mathbf{x}}_i.\end{aligned}$$

That means that for any solution at all $\mathbf{k}'\tilde{\mathbf{x}} = \mathbf{k}'\mathbf{G}\mathbf{y}$ if $\mathbf{k}'\mathbf{H} = \mathbf{k}'$.

To prove the necessity part of the theorem, choose $\mathbf{z}^* = \mathbf{0}$ in $\mathbf{x}^*$. Then

$$\begin{aligned}\mathbf{k}'\mathbf{x}^* = \mathbf{k}'\mathbf{G}\mathbf{y} &= \mathbf{k}'\sum_{i=1}^{q-r+1} \lambda_i \tilde{\mathbf{x}}_i = \mathbf{k}'\sum_{i=1}^{q-r+1} \lambda_i[\mathbf{G}\mathbf{y} + (\mathbf{H} - \mathbf{I})\mathbf{z}_i] \\ &= \mathbf{k}'\mathbf{G}\mathbf{y}\left(\sum_{i=1}^{q-r+1} \lambda_i\right) + \mathbf{k}'\sum_{i=1}^{q-r+1} \lambda_i(\mathbf{H} - \mathbf{I})\mathbf{z}_i \\ &= \mathbf{k}'\mathbf{G}\mathbf{y} + \mathbf{k}'\sum_{i=1}^{q-r+1} \lambda_i(\mathbf{H} - \mathbf{I})\mathbf{z}_i.\end{aligned}$$

Hence, $\mathbf{k}'\sum_{i=1}^{q-r+1}\lambda_i(\mathbf{H}-\mathbf{I})\mathbf{z}_i = 0$. However, the λ_i are not all zero and the $(\mathbf{H}-\mathbf{I})\mathbf{z}_i$ are LIN. Therefore, this last equation can be true only if $\mathbf{k}'(\mathbf{H}-\mathbf{I}) = 0$, that is, $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. Hence, for any solution $\mathbf{x}^*$, $\mathbf{k}'\mathbf{x}^* = \mathbf{k}'\mathbf{G}\mathbf{y}$ if and only if $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. This proves the theorem conclusively. ■

Example 8 Illustration of the Invariance Principle

In deriving (14) in Example 6, we have that

$$\mathbf{H} = \mathbf{GA} = \begin{bmatrix} 1 & 0 & -1 & -29 \\ 0 & 1 & 2 & 47 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and for

$$\mathbf{k}' = \begin{bmatrix} 3 & 2 & 1 & 7 \end{bmatrix}, \tag{20}$$

it will be found that $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. Therefore, $\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to whatever solution is used for $\tilde{\mathbf{x}}$. Thus, from (15) and (16)

$$\mathbf{k}'\tilde{\mathbf{x}}_1 = 3(6) + 2(-8) + 1(0) + 7(0) = 2$$

and

$$\mathbf{k}'\tilde{\mathbf{x}}_2 = 3(-51) + 2(84) + 1(1) + 7(-2) = 2.$$

In general, from (14)

$$\mathbf{k}'\tilde{\mathbf{x}} = 3(6 - z_3 - 29z_4) + 2(-8 + 2z_3 + 47z_4) + 1(-z_3) + 7(-z_4) = 2.$$

Likewise, $\mathbf{k}'\dot{\mathbf{x}}$ has the same value. From (17)

$$\mathbf{k}'\dot{\mathbf{x}} = 3(-\dot{z}_1) + 2(4 + 2\dot{z}_1 - 11\dot{z}_4) + 1(-6 - \dot{z}_1 + 29\dot{z}_4) + 7(-\dot{z}_4) = 2.$$

There are of course many values of $\mathbf{k}'$ that satisfy $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. For each of these $\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to the choice of $\tilde{\mathbf{x}}$. For two such vectors $\mathbf{k}_1'$ and $\mathbf{k}_2'$ say $\mathbf{k}_1'\tilde{\mathbf{x}}$ and $\mathbf{k}_2'\tilde{\mathbf{x}}$ are different but each has a value that is the same for all values of $\tilde{\mathbf{x}}$. Thus, in the example $\mathbf{k}_1'\mathbf{H} = \mathbf{k}_1'$, where

$$\mathbf{k}_1' = \begin{bmatrix} 1 & 2 & 3 & 65 \end{bmatrix}$$

is different from (20) and

$$\mathbf{k}_1'\tilde{\mathbf{x}}_1 = 1(6) + 2(-8) + 3(0) + 65(0) = -10$$

is different from $\mathbf{k}'\tilde{\mathbf{x}}$ for $\mathbf{k}'$ of (20). However, for every $\tilde{\mathbf{x}}$, $\mathbf{k}_1'\tilde{\mathbf{x}}_1 = -10$. □

It was shown in Theorem 6 that the invariance of $\mathbf{k}'\tilde{\mathbf{x}}$ to $\tilde{\mathbf{x}}$ holds for any $\mathbf{k}'$ provided that $\mathbf{k}'\mathbf{H} = \mathbf{k}'$. Two corollaries of the theorem follow.

Corollary 6.1 The linear combination $\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to $\tilde{\mathbf{x}}$ for $\mathbf{k}'$ of the form $\mathbf{k}' = \mathbf{w}'\mathbf{H}$ for arbitrary $\mathbf{w}'$.

Proof. We have that $\mathbf{k}'\mathbf{H} = \mathbf{w}'\mathbf{H}^2 = \mathbf{w}'\mathbf{GAGA} = \mathbf{w}'\mathbf{GA} = \mathbf{w}'\mathbf{H} = \mathbf{k}'$. ∎

Corollary 6.2 There are only r LIN vectors $\mathbf{k}'$ for which $\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to $\tilde{\mathbf{x}}$.

Proof. Since $r(\mathbf{H}) = r$, there are in $\mathbf{k}' = \mathbf{w}'\mathbf{H}$ of order q exactly $q - r$ elements that are linear combinations of the other r. Therefore, for arbitrary vectors $\mathbf{w}'$ there are only r LIN vectors $\mathbf{k}' = \mathbf{w}'\mathbf{H}$. ∎

We will return to this point in Chapter 5 when we discuss estimable functions.

The concept of generalized inverse has now been defined and its use in solving linear equations explained. Next, we briefly discuss the generalized inverse itself, its various definitions and some of its properties. Extensive review of generalized inverses and their applications is to be found in Boullion and Odell (1968) and the approximately 350 references there. A more recent reference on generalized inverses is Ben-Israel and Greville (2003).

3. THE PENROSE INVERSE

Penrose (1955) in extending the work of Moore (1920), shows that for every matrix $\mathbf{A}$, there is a unique matrix $\mathbf{K}$ which satisfies the following conditions:

$$\begin{aligned} &\mathbf{AKA} = \mathbf{A} && \text{(i)} \\ &\mathbf{KAK} = \mathbf{K} && \text{(ii)} \\ &(\mathbf{KA})' = \mathbf{KA} && \text{(iii)} \\ &(\mathbf{AK})' = \mathbf{AK} && \text{(iv)} \end{aligned} \tag{21}$$

Such generalized inverses $\mathbf{K}$ will be referred to as Moore–Penrose inverses. We will show how to find them and prove that every matrix has a unique Moore–Penrose inverse.

Condition (i) states that $\mathbf{K}$ is a generalized inverse of $\mathbf{A}$. Condition (ii) states that $\mathbf{A}$ is a generalized inverse of $\mathbf{K}$. In Section 4, we will give an example to show that in general, Condition (i) does not imply condition (ii). Conditions (iii) and (iv) state that $\mathbf{KA}$ and $\mathbf{AK}$, respectively, are symmetric matrices. There are generalized inverses that satisfy one or more of conditions (ii), (iii), and (iv) but not all of them. We will give examples of such generalized inverses in Section 4.

In Section 2, we showed how to obtain a generalized inverse using the singular value decomposition. These generalized inverses satisfy all four of the above conditions and, as a result, are Moore–Penrose inverses. Although a matrix has infinitely many generalized inverses it has only one Moore–Penrose inverse. We show this below.

Theorem 7 Let $\mathbf{A}$ be a matrix with singular value decomposition $\mathbf{S}'\Lambda^{1/2}\mathbf{U}'$. Then the generalized inverse $\mathbf{K} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{S}$ is the Moore–Penrose inverse.

Proof. We have already shown that $\mathbf{K}$ is in fact a generalized inverse. To establish the second Penrose condition we have

$$\mathbf{KAK} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{SS}'\boldsymbol{\Lambda}^{1/2}\mathbf{U}'\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{S} = \mathbf{K}.$$

Now

$$\mathbf{KA} = \mathbf{U}\Lambda^{-1/2}\mathbf{SS}'\Lambda^{1/2}\mathbf{U}' = \mathbf{UU}'.$$

and

$$\mathbf{AK} = \mathbf{S}'\boldsymbol{\Lambda}^{1/2}\mathbf{U}'\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{S} = \mathbf{S}'\mathbf{S}.$$

Since $\mathbf{UU}'$ and $\mathbf{SS}'$ are symmetric matrices, conditions (iii) and (iv) are established. ∎

The generalized inverse in Example 5 is the Moore–Penrose inverse of $\mathbf{A}$ there.

We have thus established the existence of the Moore–Penrose inverse. We now show that it is unique.

Theorem 8 The Moore–Penrose inverse is unique.

Proof. The proof consists of showing that for a given matrix, there can be only one matrix that satisfies the four conditions. First, from condition (i) and (iii)

$\mathbf{A} = \mathbf{AKA} = \mathbf{A}(\mathbf{KA})' = \mathbf{AA}'\mathbf{K}'$ and by transposition

$$\mathbf{KAA}' = \mathbf{A}'. \tag{22}$$

Also if $\mathbf{AA}'\mathbf{K}' = \mathbf{A}$, then $\mathbf{KAA}'\mathbf{K}' = \mathbf{KA}(\mathbf{KA})' = \mathbf{KA}$ so that $\mathbf{KA}$ is symmetric. Also $\mathbf{AKA} = \mathbf{A}(\mathbf{KA})' = \mathbf{AA}'\mathbf{K}' = \mathbf{A}$. Thus (22) is equivalent to (i) and (iii).

Likewise, from condition (ii) and (iv), we can show in a similar way that an equivalent identity is

$$\mathbf{KK}'\mathbf{A}' = \mathbf{K}. \tag{23}$$

Suppose that $\mathbf{K}$ is not unique. Assume some other matrix $\mathbf{M}$ satisfies the Penrose conditions. From conditions (i) and (iv), we have

$$\mathbf{AA'M} = \mathbf{A'} \tag{24}$$

and from conditions (ii) and (iii)

$$\mathbf{A'M'M} = \mathbf{M}. \tag{25}$$

We then have that using (22)–(25),

$$\mathbf{K} = \mathbf{KK'A'} = \mathbf{KK'A'AM} = \mathbf{KAM} = \mathbf{KAA'M'M} = \mathbf{A'MM'} = \mathbf{M}.$$

This establishes uniqueness. ■

We now give another method for finding the Moore–Penrose inverse based on the Cayley–Hamilton Theorem (see, for example, Searle (1966), C. R. Rao (1973), and Gruber (2014)). The Cayley–Hamilton theorem states that a square matrix satisfies its characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. To show this, we need two lemmas.

Lemma 2 If the matrix $\mathbf{X'X} = 0$, then $\mathbf{X} = 0$.

Proof. If the matrix $\mathbf{X'X} = 0$, then the sums of squares of the elements of each row are zero so that the elements themselves are zero. ■

Lemma 3 The identity $\mathbf{PX'X} = \mathbf{QX'X}$ implies that $\mathbf{PX'} = \mathbf{QX'}$.

Proof. Apply Lemma 2 to

$$(\mathbf{PX'X} - \mathbf{QX'X})(\mathbf{P} - \mathbf{Q}) = (\mathbf{PX'} - \mathbf{QX'})(\mathbf{PX'} - \mathbf{QX'})' = 0.$$

We will give an alternative proof that uses the singular value decomposition of $\mathbf{X} = \mathbf{S'}\boldsymbol{\Lambda}^{1/2}\mathbf{U'}$. We have that $\mathbf{PX'X} = \mathbf{QX'X}$ implies that $\mathbf{PU}\Lambda^{1/2}\mathbf{SS'}\Lambda^{1/2}\mathbf{U'} = \mathbf{QU}\Lambda^{1/2}\mathbf{SS'}\Lambda^{1/2}\mathbf{U'}$.

Multiply both sides of this equation by $\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{S}$ and obtain $\mathbf{PU}\Lambda^{1/2}\mathbf{SS'S} = \mathbf{QU}\Lambda^{1/2}\mathbf{SS'S}$. Since $\mathbf{SS'} = \mathbf{I}$, we have

$$\mathbf{PU}\Lambda^{1/2}\mathbf{S} = \mathbf{QU}\Lambda^{1/2}\mathbf{S} \text{ or } \mathbf{PX'} = \mathbf{QX'}.$$

■

We now assume that

$$\mathbf{K} = \mathbf{TA'} \tag{26}$$

for some matrix **T**. Then (22) is satisfied if **T** satisfies

$$\mathbf{TA'AA'} = \mathbf{A'}; \tag{27}$$

and since satisfaction of (22) implies satisfaction of conditions (i) and (iii). Thus,

$\mathbf{AKA} = \mathbf{A}$ and $\mathbf{A'K'A'} = \mathbf{A'}$. As a result, $\mathbf{TA'K'A'} = \mathbf{TA'}$, or from (26), we get (23).

However, (23) is equivalent to Penrose conditions (ii) and (iv) so **K** as defined in (26) for **T** that satisfies (27).

We now derive a suitable **T.** Notice that the matrix $\mathbf{A'A}$ and all of its powers are square. By the Cayley–Hamilton Theorem, for some integer t, there exists a series of scalars $\lambda_1, \lambda_2, \dots, \lambda_t$ not all zero, such that

$$\lambda_1\mathbf{A'A} + \lambda_2(\mathbf{A'A})^2 + \cdots + \lambda_t(\mathbf{A'A})^t = 0.$$

If λ_r is the first λ in this identity that is non-zero then **T** is defined as

$$\mathbf{T} = (-1/\lambda_r)[\lambda_{r+1}\mathbf{I} + \lambda_{r+2}(\mathbf{A'A}) + \cdots + \lambda_t(\mathbf{A'A})^{t-r-1}]. \tag{28}$$

To show that this satisfies (27) note that by direct multiplication

$$\begin{aligned}\mathbf{T}(\mathbf{A'A})^{r+1} &= (-1/\lambda_r)[\lambda_{r+1}(\mathbf{A'A})^{r+1} + \lambda_{r+2}(\mathbf{A'A})^{r+2} + \cdots \lambda_t(\mathbf{A'A})^t] \\ &= (-1/\lambda_r)[-\lambda_1\mathbf{A'A} - \lambda_2(\mathbf{A'A})^2 - \cdots \lambda_r(\mathbf{A'A})^r].\end{aligned}$$

Since by definition λ_r is the first non-zero λ in the series $\lambda_1, \lambda_2, \dots$, the above reduces to

$$\mathbf{T}(\mathbf{A'A})^{r+1} = (\mathbf{A'A})^r. \tag{29}$$

Repeated use of Lemma 3 reduces this to (27). Thus, $\mathbf{K} = \mathbf{TA'}$ with **T** as defined in (28) satisfies (27) and hence is the unique generalized inverse satisfying all four of the Penrose conditions.

Example 9 Finding a Moore–Penrose Inverse using the Cayley–Hamilton Theorem
For

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \\ -1 & 0 & -2 \\ 1 & 2 & 0 \end{bmatrix}, \text{ we have } \mathbf{A'A} = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 5 & -1 \\ 4 & -1 & 9 \end{bmatrix}.$$

Finding the characteristic equation $66\lambda - 17\lambda^2 + \lambda^3 = 0$ and employing the Cayley–Hamilton theorem, we have

$$66(\mathbf{A'A}) - 17(\mathbf{A'A})^2 + (\mathbf{A'A})^3 = 0.$$

Then,

$$\mathbf{T} = (-1/66)(-17\mathbf{I} + \mathbf{A}'\mathbf{A}) = (1/66)\begin{bmatrix} 14 & -2 & -4 \\ -2 & 12 & 1 \\ -4 & 1 & 8 \end{bmatrix}$$

and $\mathbf{K} = \mathbf{TA}' = (1/66)\begin{bmatrix} 6 & -2 & -6 & 10 \\ 0 & -11 & 0 & 22 \\ 12 & 7 & -12 & -2 \end{bmatrix}$ is the Penrose inverse of $\mathbf{A}$ satisfying 21. □

Graybill et al. (1966) suggests an alternative procedure for deriving $\mathbf{K}$. Their method is to find X and $\mathbf{Y}$ such that

$$\mathbf{AA}'\mathbf{X}' = \mathbf{A} \quad \text{and} \quad \mathbf{A}'\mathbf{AY} = \mathbf{A}' \tag{30}$$

and then

$$\mathbf{K} = \mathbf{XAY}. \tag{31}$$

Proof that $\mathbf{K}$ satisfies all four Penrose axioms depends on using (30) and Lemma 3 to show that $\mathbf{AXA} = \mathbf{A} = \mathbf{AYA}$. (See Exercise 28.)

4. OTHER DEFINITIONS

It is clear that the Penrose inverse $\mathbf{K}$ is not easy to compute, especially when $\mathbf{A}$ has many columns or irrational eigenvalues because either finding the singular value decomposition or using the Cayley–Hamilton theorem can be quite tedious. As has already been shown, only the first Penrose condition needs to be satisfied to have a matrix useful for solving linear equations. Furthermore, in pursuing the topic of linear models, this is the only condition that is really needed. For this reason, a generalized inverse has been defined as any matrix that satisfies $\mathbf{AGA} = \mathbf{A}$. This definition will be retained throughout the book. Nevertheless, a variety of names will be found throughout the literature, both for $\mathbf{G}$ and for other matrices satisfying fewer than all four of the Penrose conditions. There are five such possibilities as detailed in Table 1.1.

In the notation of Table 1.1 $\mathbf{A}^{(\mathrm{g})} = \mathbf{G}$, the generalized inverse already defined and discussed, and $\mathbf{A}^{(\mathrm{p})} = \mathbf{K}$, the Moore–Penrose inverse. This has also been called the pseudo inverse and the p-inverse by various authors. The Software package Mathematica computes the Moore–Penrose inverse of $\mathbf{A}$ in response to the input PseudoInverse[A]. The suggested definition of normalized generalized inverse in Table 1.1 is not universally accepted. As given there it is used by Urquhart (1968), whereas Goldman and Zelen (1964), call it a "weak" generalized inverse. An example of such a matrix is a left inverse $\mathbf{L}$ such that $\mathbf{LA} = \mathbf{I}$. Rohde (1966) has also used the description "normalized" (we use reflexive least square) for a matrix satisfying

TABLE 1.1 Suggested Names for Matrices Satisfying Some or All of the Penrose Conditions

Conditions Satisfied (Eq. 21)	Name of Matrix	Symbol
i	Generalized inverse	$\mathbf{A}^{(g)}$
i and ii	Reflexive generalized inverse	$\mathbf{A}^{(r)}$
i and iii	Mininum norm generalized inverse	$\mathbf{A}^{(mn)}$
i and iv	Least-square generalized inverse	$\mathbf{A}^{(ls)}$
i, ii, and iii	Normalized generalized inverse	$\mathbf{A}^{(n)}$
i, ii, and iv	Reflexive least square Generalized inverse	$\mathbf{A}^{(rls)}$
i, ii, iii, and iv	Moore–Penrose inverse	$\mathbf{A}^{(p)}$

conditions (i), (ii), and (iv). An example of this kind of matrix is a right inverse $\mathbf{R}$ for which $\mathbf{AR} = \mathbf{I}$.

The generalized inverses obtained in Section 1 by diagonalization or the algorithm are reflexive. See Exercise 27.

Let $\mathbf{x} = \mathbf{Gy}$ be a solution to $\mathbf{Ax} = \mathbf{y}$. The minimum norm generalized inverse is such that $\min_{\mathbf{Ax=y}} \|\mathbf{x}\| = \|\mathbf{Gy}\|$. Such a generalized inverse satisfies Penrose conditions (i) and (iii). The least-square generalized inverse is the one that yields the solution $\tilde{\mathbf{x}}$ such that $\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{y}\| = \inf_x \|\mathbf{Ax} - \mathbf{y}\|$. It satisfies Penrose conditions (i) and (iv). Proofs of these results are available in Gruber (2014), and Rao and Mitra (1971).

The following relationships can be established between the generalized inverses.

$$\begin{aligned}
\mathbf{A}^{(r)} &= \mathbf{A}^{(g)}\mathbf{A}\mathbf{A}^{(g)} \\
\mathbf{A}^{(n)} &= \mathbf{A}'(\mathbf{A}\mathbf{A}')^{(g)} \\
\mathbf{A}^{(rls)} &= (\mathbf{A}'\mathbf{A})^{(g)}\mathbf{A}' \\
\mathbf{A}^{(p)} &= \mathbf{A}^{(n)}\mathbf{A}\mathbf{A}^{(rls)}
\end{aligned} \tag{32}$$

Some general conditions for generalized inverses to be reflexive, minimum norm or least square are developed in Gruber (2014).

Example 10 Finding Different Kinds of Generalized Inverses

As in Example 9,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \\ -1 & 0 & -2 \\ 1 & 2 & 0 \end{bmatrix}, \mathbf{A}'\mathbf{A} = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 5 & -1 \\ 4 & -1 & 9 \end{bmatrix}, \text{ and } \mathbf{A}\mathbf{A}' = \begin{bmatrix} 5 & 2 & -5 & 1 \\ 2 & 2 & -2 & -2 \\ -5 & -2 & 5 & -1 \\ 1 & -2 & -1 & 5 \end{bmatrix}.$$

These three matrices have rank 2. Using the algorithm in Part 1,

$$\mathbf{A}^{(g)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since this is a reflexive generalized inverse $\mathbf{A}^{(r)} = \mathbf{A}^{(g)}$. Now,

$$(\mathbf{AA}')^{(g)} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & \frac{5}{6} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{A}^{(n)} = \mathbf{A}'(\mathbf{AA}')^{(g)} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ \frac{1}{3} & -\frac{5}{6} & 0 & 0 \\ \frac{1}{3} & \frac{1}{6} & 0 & 0 \end{bmatrix}.$$

Furthermore,

$$(\mathbf{A}'\mathbf{A})^{(g)} = \begin{bmatrix} \frac{5}{11} & -\frac{2}{11} & 0 \\ -\frac{2}{11} & \frac{3}{11} & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{A}^{(rls)} = (\mathbf{A}'\mathbf{A})^{(g)}\mathbf{A}' = \begin{bmatrix} \frac{5}{11} & \frac{2}{11} & -\frac{5}{11} & \frac{1}{11} \\ -\frac{2}{11} & -\frac{3}{11} & \frac{2}{11} & \frac{4}{11} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then

$$\mathbf{A}^{(p)} = \mathbf{A}^{(n)}\mathbf{A}\mathbf{A}^{(rls)} = (1/66)\begin{bmatrix} 6 & -2 & -6 & 10 \\ 0 & -11 & 0 & 22 \\ 12 & 7 & -12 & -2 \end{bmatrix}.$$

□

5. SYMMETRIC MATRICES

The study of linear models frequently leads to equations of the form $\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{y}$ that have to be solved for $\hat{\mathbf{b}}$. Special attention is given therefore to the properties of a generalized inverse of the symmetric matrix $\mathbf{X}'\mathbf{X}$.

a. Properties of a Generalized Inverse

The facts summarized in Theorem 9 below will be useful. We will denote the Moore–Penrose inverse by $(\mathbf{X}'\mathbf{X})^+$ and any generalized inverse by $(\mathbf{X}'\mathbf{X})^-$

Theorem 9 Assume that the singular value decomposition of $\mathbf{X} = \mathbf{S}'\mathbf{\Lambda}^{1/2}\mathbf{U}'$. Then

(i) $\mathbf{X}'\mathbf{X} = \mathbf{U}\Lambda\mathbf{U}'$ and $(\mathbf{X}'\mathbf{X})^+ = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}'$.
(ii) For any generalized inverse of $\mathbf{X}'\mathbf{X}$, $\mathbf{U}'(\mathbf{X}'\mathbf{X})^-\mathbf{U} = \mathbf{\Lambda}^{-1}$ and therefore $(\mathbf{X}'\mathbf{X})^+ = \mathbf{UU}'(\mathbf{X}'\mathbf{X})^-\mathbf{UU}'$.

(iii) Any generalized inverse $\mathbf{G}$ of $\mathbf{X}'\mathbf{X}$ may be written in terms of the Moore–Penrose inverse as follows

$$\begin{aligned}\mathbf{G} &= (\mathbf{X}'\mathbf{X})^{+} + \mathbf{V}\mathbf{C}_1\mathbf{U}' + \mathbf{U}\mathbf{C}_1'\mathbf{V}' + \mathbf{V}\mathbf{C}_2\mathbf{V}' \\ &= \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}\begin{bmatrix}\Lambda^{-1} & \mathbf{C}_1 \\ \mathbf{C}_1' & \mathbf{C}_2\end{bmatrix}\begin{bmatrix}\mathbf{U} \\ \mathbf{V}\end{bmatrix},\end{aligned}$$

where $\mathbf{C}_1 = \mathbf{V}'\mathbf{G}\mathbf{U}$ and $\mathbf{C}_2 = \mathbf{V}'\mathbf{G}\mathbf{V}$

Proof. For (i) $\mathbf{X}'\mathbf{X} = \mathbf{U}\Lambda^{1/2}\mathbf{S}\mathbf{S}'\Lambda^{1/2}\mathbf{U}' = \mathbf{U}\Lambda\mathbf{U}'$ because $\mathbf{S}\mathbf{S}' = \mathbf{I}$. The expression $\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}'$ can be shown to satisfy the Penrose conditions.

For (ii) we have that $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$.

Then this implies that

$$\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}' = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'. \tag{33}$$

Multiply both sides of equation (33) on the left by $\boldsymbol{\Lambda}^{-1}\mathbf{U}'$ and on the right by $\mathbf{U}\boldsymbol{\Lambda}^{-1}$. The result follows.

To establish (iii), notice that

$$\begin{aligned}\mathbf{G} &= (\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}')\mathbf{G}(\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}') \\ &= \mathbf{U}\mathbf{U}'\mathbf{G}\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}'\mathbf{G}\mathbf{U}\mathbf{U}' + \mathbf{U}\mathbf{U}'\mathbf{G}\mathbf{V}\mathbf{V}' + \mathbf{V}\mathbf{V}'\mathbf{G}\mathbf{V}\mathbf{V}' \\ &= (\mathbf{X}'\mathbf{X})^{+} + \mathbf{V}\mathbf{C}_1\mathbf{U}' + \mathbf{U}\mathbf{C}_1'\mathbf{V}' + \mathbf{V}\mathbf{C}_2\mathbf{V}' \\ &= \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}^{-1} & \mathbf{C}_1 \\ \mathbf{C}_1' & \mathbf{C}_2\end{bmatrix}\begin{bmatrix}\mathbf{U} \\ \mathbf{V}\end{bmatrix}.\end{aligned}$$

Theorem 10 below gives some more useful properties of generalized inverses of $\mathbf{X}'\mathbf{X}$. ■

Theorem 10 When $\mathbf{G}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$ then

(i) $\mathbf{G}'$ is also a generalized inverse of $\mathbf{X}'\mathbf{X}$;
(ii) $\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}$; that is, $\mathbf{G}\mathbf{X}'$ is a generalized inverse of $\mathbf{X}$;
(iii) $\mathbf{X}\mathbf{G}\mathbf{X}'$ is invariant to $\mathbf{G}$;
(iv) $\mathbf{X}\mathbf{G}\mathbf{X}'$ is symmetric whether G is or not.

Proof.

(i) By definition, $\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$. Transposition yields $\mathbf{X}'\mathbf{X}\mathbf{G}'\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$.
(ii) Observe that $\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{S}'\boldsymbol{\Lambda}^{1/2}\mathbf{U}'\mathbf{G}\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}' = \mathbf{S}'\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\mathbf{U}' = \mathbf{S}'\boldsymbol{\Lambda}^{1/2}\mathbf{U}' = \mathbf{X}$.

The result may also be obtained by application of Lemma 3.

(iii) Notice that $\mathbf{XGX'} = \mathbf{S'\Lambda^{1/2}U'GU\Lambda^{1/2}S} = \mathbf{S'\Lambda^{1/2}\Lambda^{-1}\Lambda^{1/2}S} = \mathbf{S'S}$.

(iv) If $\mathbf{M}$ is a symmetric generalized inverse of $\mathbf{X'X}$ then $\mathbf{XMX'}$ is symmetric. (For example, the Moore–Penrose inverse of $\mathbf{X'X}$ is symmetric.) From (iii) $\mathbf{XGX'} = \mathbf{XMX'}$ and is thus, symmetric whether or not $\mathbf{G}$ is.

■

Corollary 10.1 Applying part (i) of Theorem 10 to the other parts shows that

$$\mathbf{XGX'X} = \mathbf{X}, \mathbf{X'XGX'} = \mathbf{X'} \text{ and } \mathbf{X'XG'X} = \mathbf{X'}.$$

Furthermore,

$$\mathbf{XG'X'} = \mathbf{XGX'} \text{ and } \mathbf{XG'X'} \text{ is symmetric.}$$

It is to be emphasized that not all generalized inverses of a symmetric matrix are symmetric. This is illustrated in Example 11 below.

Example 11 The Generalized Inverse of a Symmetric Matrix Need not be Symmetric

We can demonstrate this by applying the algorithm at the end of Section 1 to the symmetric matrix using the sub-matrix from the first two columns of the first and third rows

$$\mathbf{A}_2 = \begin{bmatrix} 2 & 2 & 6 \\ 2 & 3 & 8 \\ 6 & 8 & 22 \end{bmatrix}$$

to obtain the non-symmetric generalized inverse

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

□

Theorem 10 and Corollary 10.1 very largely enable us to avoid difficulties that this lack of symmetry of generalized inverses of $\mathbf{X'X}$ might otherwise appear to involve. For example, if $\mathbf{G}$ is a generalized inverse of $\mathbf{X'X}$ and $\mathbf{P}$ is some other matrix,

$$(\mathbf{PXGX'})' = \mathbf{XG'X'P'} = \mathbf{XGX'P'}$$

not because $\mathbf{G}$ is symmetric (which in general is not) but because $\mathbf{XGX'}$ is symmetric.

□

Example 12 Illustration of Symmetry of $\mathbf{XGX'}$

If

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 0 & 1 & 2 \end{bmatrix},$$

then $\mathbf{X'X} = \mathbf{A}_2$ from Example 11. Then

$$\mathbf{XGX'} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 3 & 3 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

and

$$\mathbf{XG'X'} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 3 & 3 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

□

b. Two More Generalized Inverses of X′X

In addition to the methods studied already, two other methods discussed by John (1964) are sometimes pertinent to linear models. They depend on the ordinary inverse of a non-singular matrix:

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{X'X} & \mathbf{H'} \\ \mathbf{H} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} = \mathbf{0} \end{bmatrix}. \tag{34}$$

The matrix $\mathbf{H}$ being used here is not the matrix $\mathbf{H} = \mathbf{GA}$ used earlier. It is being used to be consistent with John's notation. The matrix $\mathbf{X'X}$ is of order p and rank $p - m$. The matrix $\mathbf{H}$ is any matrix of order $m \times p$. It is of full row rank and its rows also LIN of those of $\mathbf{X'X}$. In other words, the rows of $\mathbf{H}$ cannot be a linear combination of rows of $\mathbf{X'X}$. (The existence of such a matrix is assured by considering m vectors of order p that are LIN of any set of $p - m$ LIN rows of $\mathbf{X'X}$. Furthermore, if these rows constitute $\mathbf{H}$ in such a way that the m LIN rows of $\mathbf{H}$ correspond in $\mathbf{S}$ to the m rows of $\mathbf{X'X}$ that are linear combinations of the set of $p - m$ rows then $\mathbf{S}^{-1}$ of (34) exists.) With (34) existing the two matrices

$$\mathbf{B}_{11} \text{ and } (\mathbf{X'X} + \mathbf{H'H})^{-1} \text{ are generalized inverses of } \mathbf{X'X}. \tag{35}$$

Three useful lemmas help in establishing these results.

Lemma 4 The matrix $\mathbf{T} = \begin{bmatrix} \mathbf{I}_r & \mathbf{U} \end{bmatrix}$ has rank r for any matrix $\mathbf{U}$ of r rows.

Proof. Elementary operations carried out on $\mathbf{T}$ to find its rank will operate on $\mathbf{I}_r$. None of these rows or columns can be made null by such operations. Therefore, $r(\mathbf{T})$ is not less than r. Consequently $r(\mathbf{T}) = r$. ■

Lemma 5 If $\mathbf{X}_{N\times p}$ has rank $p - m$ for $m > 0$, then there exists a matrix $\mathbf{D}_{p\times m}$ such that $\mathbf{XD} = \mathbf{0}$ and $r(\mathbf{D}) = m$.

Proof. Let $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$ where $\mathbf{X}_1$ is $N \times (p - m)$ of full column rank. Then the columns of $\mathbf{X}_2$ are linear combinations of the columns of $\mathbf{X}_1$ and so for some matrix $\mathbf{C}$, of order $(p - m) \times m$, the sub-matrices of $\mathbf{X}$ satisfy $\mathbf{X}_2 = \mathbf{X}_1\mathbf{C}$. Let $\mathbf{D}' = \begin{bmatrix}-\mathbf{C}' & \mathbf{I}_m\end{bmatrix}$. By Lemma 4 $\mathbf{D}'$ has rank m. We then have $\mathbf{XD} = \mathbf{0}$ and $r(\mathbf{D}) = m$. The Lemma is thus proved because a matrix $\mathbf{D}$ exists. ■

Lemma 6 For $\mathbf{X}$ and $\mathbf{D}$ of Lemma 5 and $\mathbf{H}$ of order $m \times p$ with full-row rank, $\mathbf{HD}$ has full-row rank if and only if the rows of $\mathbf{H}$ are LIN of those of $\mathbf{X}$.

Proof. (i) Given $r(\mathbf{HD}) = m$, assume that the rows of $\mathbf{H}$ depend on those of $\mathbf{X}$ (are not LIN of $\mathbf{X}$). Then, $\mathbf{H} = \mathbf{KX}$ for some $\mathbf{K}$, and $\mathbf{HD} = \mathbf{KXD} = \mathbf{0}$. Therefore, the assumption is false and the rows of $\mathbf{H}$ are LIN of those of $\mathbf{X}$.

(ii) Given that the rows of $\mathbf{H}$ are LIN of those of $\mathbf{X}$, the matrix $\begin{bmatrix}\mathbf{X}\\ \mathbf{R}\end{bmatrix}$, of order $(N + m) \times p$ has full column rank. Therefore, it has a left inverse $\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}$, say (Section 5.13 of Searle (1966)), and so $\mathbf{UX} + \mathbf{VH} = \mathbf{I}$, that is, $\mathbf{UXD} + \mathbf{VHD} = \mathbf{D}$; or $\mathbf{VHD} = \mathbf{D}$ using Lemma 5. However, $r(\mathbf{D}_{p\times m}) = m$ and $\mathbf{D}$ has a left inverse, $\mathbf{E}$, say, and so $\mathbf{EVHD} = \mathbf{I}_m$. Therefore, $r(\mathbf{HD}) \geq m$ and so because $\mathbf{HD}$ is $m \times m$, $r(\mathbf{HD}) = m$, and the lemma is proved. ■

Proof of (35). First it is necessary to show that in (34), $\mathbf{B}_{22} = \mathbf{0}$. From (34), we have that

$$\mathbf{X}'\mathbf{X}\mathbf{B}_{11} + \mathbf{H}'\mathbf{B}_{21} = \mathbf{I} \quad \text{and} \quad \mathbf{X}'\mathbf{X}\mathbf{B}_{12} + \mathbf{H}'\mathbf{B}_{22} = \mathbf{0} \tag{36}$$

$$\mathbf{H}\mathbf{B}_{11} = \mathbf{0} \quad \text{and} \quad \mathbf{H}\mathbf{B}_{12} = \mathbf{I}. \tag{37}$$

Pre-multiplying (36) by $\mathbf{D}'$ and using Lemmas 5 and 6 leads to

$$\mathbf{B}_{21} = (\mathbf{D}'\mathbf{H}')^{-1}\mathbf{D}' \quad \text{and} \quad \mathbf{B}_{22} = \mathbf{0}. \tag{38}$$

Then from (36) and (38),

$$\mathbf{X}'\mathbf{X}\mathbf{B}_{11} = \mathbf{I} - \mathbf{H}'(\mathbf{D}'\mathbf{H}')^{-1}\mathbf{D}'. \tag{39}$$

Post-multiplication of (39) by $\mathbf{X}'\mathbf{X}$ and application of Lemma 5 shows that $\mathbf{B}_{11}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$. Furthermore, using (37), (39), and Lemmas 5 and 6 gives

$$(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})[\mathbf{B}_{11} + \mathbf{D}(\mathbf{D}'\mathbf{H}'\mathbf{H}\mathbf{D})^{-1}\mathbf{D}'] = \mathbf{I}. \tag{40}$$

From (40),

$$(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1} = \mathbf{B}_{11} + \mathbf{D}(\mathbf{D}'\mathbf{H}'\mathbf{H}\mathbf{D})^{-1}\mathbf{D}'. \tag{41}$$

Since from Lemma 5, $\mathbf{D}$ is such that $\mathbf{XD} = \mathbf{0}$ we have that

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}\mathbf{B}_{11}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$$

since $\mathbf{B}_{11}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$ and $(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$. This completes the proof. ■

It can be shown that $\mathbf{B}_{11}$ satisfies the second of Penrose conditions and is thus a reflexive generalized inverse of $\mathbf{X}'\mathbf{X}$. However, $(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}$ only satisfies the first Penrose condition. Neither generalized inverse satisfies conditions (iii) or (iv).

John (1964) refers to Graybill (1961, p. 292) and to Kempthorne (1952, p. 79) in discussing $\mathbf{B}_{11}$ and to Plackett (1960, p. 41) and Scheffe (1959, p. 19) in discussing $(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}$, in terms of defining generalized inverses of $\mathbf{X}'\mathbf{X}$ as being matrices $\mathbf{G}$ for which $\mathbf{b} = \mathbf{G}\mathbf{X}'\mathbf{y}$ is a solution of $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$. By Theorem 1, they then satisfy condition (i), as has just been shown. Rayner and Pringle (1967) also discuss these results, indicating that $\mathbf{D}$ of the previous discussion may be taken as $(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$. This, as Chipman (1964) shows, means that $\mathbf{HD} = \mathbf{I}$ and so (39) becomes

$$\mathbf{X}'\mathbf{X}\mathbf{B}_{11} = \mathbf{I} - \mathbf{H}'\mathbf{H}(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}, \tag{42}$$

a simplified form of Rayner and Pringle's equation (7). The relationship between the two generalized inverses of $\mathbf{X}'\mathbf{X}$ shown in (35) is therefore that indicated in (42). Also note that Lemma 6 is equivalent to Theorem 3 of Scheffe (1959, p. 17).

6. ARBITRARINESS IN A GENERALIZED INVERSE

The existence of many generalized inverses $\mathbf{G}$ that satisfy $\mathbf{AGA} = \mathbf{A}$ has been emphasized. We examine here the nature of the arbitrariness of such generalized inverses as discussed in Urquhart (1969a). We need some results about the rank of the matrix. These are contained in Lemmas 7–9.

Lemma 7 A matrix of full-row rank r can be written as the product of matrices, one being of the form $\begin{bmatrix}\mathbf{I}_r & \mathbf{S}\end{bmatrix}$ for some matrix $\mathbf{S}$ of r rows.

Proof. Suppose $\mathbf{B}_{r\times q}$ has full-row rank r and contains an $r \times r$ non-singular minor, $\mathbf{M}$, say. Then, for some matrix $\mathbf{L}$ and some permutation matrix $\mathbf{Q}$ (see the paragraph just before (9)), we have $\mathbf{BQ} = \begin{bmatrix}\mathbf{M} & \mathbf{L}\end{bmatrix}$, so that

$$\mathbf{B} = \mathbf{M}\begin{bmatrix}\mathbf{I} & \mathbf{M}^{-1}\mathbf{L}\end{bmatrix}\mathbf{Q}^{-1} = \mathbf{M}\begin{bmatrix}\mathbf{I} & \mathbf{S}\end{bmatrix}\mathbf{Q}^{-1} \text{ for } \mathbf{S} = \mathbf{M}^{-1}\mathbf{L}.$$

■

Lemma 8 $\mathbf{I} + \mathbf{KK}'$ has full rank for any non-null matrix **K.**

Proof. Assume that $\mathbf{I} + \mathbf{KK}'$ does not have full rank. Then its columns are not LIN and there exists a non-null vector $\mathbf{u}$ such that

$$(\mathbf{I} + \mathbf{KK}')\mathbf{u} = \mathbf{0}, \text{ so that } \mathbf{u}'(\mathbf{I} + \mathbf{KK}')\mathbf{u} = \mathbf{u}'\mathbf{u} + \mathbf{u}'\mathbf{K}(\mathbf{u}'\mathbf{K})' = \mathbf{0}.$$

However, $\mathbf{u}'\mathbf{u}$ and $\mathbf{u}'\mathbf{K}(\mathbf{u}'\mathbf{K})'$ are both sums of squares of real numbers. Hence, their sum is zero only if their elements are zero, that is, only if $\mathbf{u} = \mathbf{0}$. This contradicts the assumption. Therefore, $\mathbf{I} + \mathbf{KK}'$ has full rank. ■

Lemma 9 When $\mathbf{B}$ has full row rank, $\mathbf{BB}'$ is non-singular.

Proof. As in Lemma 7 write $\mathbf{B} = \mathbf{M}\begin{bmatrix}\mathbf{I} & \mathbf{S}\end{bmatrix}\mathbf{Q}^{-1}$ where $\mathbf{M}^{-1}$ exists. Then because $\mathbf{Q}$ is a permutation matrix and thus orthogonal $\mathbf{BB}' = \mathbf{M}(\mathbf{I} + \mathbf{SS}')\mathbf{M}'$. By virtue of Lemma 8 and the existence of $\mathbf{M}^{-1}$, $\mathbf{BB}'$ is non-singular. ■

Corollary 9.1 When $\mathbf{B}$ has full-column rank, $\mathbf{BB}'$ is non-singular.

Proof. When $\mathbf{B}$ has full column rank $\mathbf{B}'$ has full-row rank. Now

$$\mathbf{BB}' = (\mathbf{B}'\mathbf{B})' = (\mathbf{B}'(\mathbf{B}')')'$$

From Lemma 9, $\mathbf{B}'(\mathbf{B}')'$ is non-singular and so is its transpose. ■

Consider now a matrix $\mathbf{A}_{p\times q}$ of rank r, less than both p and q. The matrix $\mathbf{A}$ contains at least one non-singular minor of order r. We will assume that this is the leading minor. There is no loss of generality in this assumption because, if it is not true, the algorithm of Section 1b will always yield a generalized inverse of $\mathbf{A}$. This generalized inverse will come from a generalized inverse of $\mathbf{B} = \mathbf{RAS}$ where $\mathbf{R}$ and $\mathbf{S}$ are permutation matrices so that $\mathbf{B}$ has a non-singular $r \times r$ leading minor. We therefore confine the discussion of inverses of $\mathbf{A}$ to the case where its leading $r \times r$ minor is non-singular. Accordingly, $\mathbf{A}$ is partitioned as

$$\mathbf{A} = \begin{bmatrix} (\mathbf{A}_{11})_{r\times r} & (\mathbf{A}_{12})_{r\times(q-r)} \\ (\mathbf{A}_{21})_{(p-r)\times r} & (\mathbf{A}_{22})_{(p-r)\times(q-r)} \end{bmatrix}. \tag{43}$$

Then since $\mathbf{A}_{11}^{-1}$ exists, $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} \end{bmatrix} \mathbf{A}_{11} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix} = \mathbf{L}\mathbf{A}_{11}\mathbf{M} \tag{44}$$

with $\mathbf{L} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} \end{bmatrix}$ and $\mathbf{M} = \begin{bmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$. From Lemma 4, $\mathbf{L}$ has full-column rank and $\mathbf{M}$ has full-row rank. Lemma 9 shows that both $(\mathbf{L}'\mathbf{L})^{-1}$ and $(\mathbf{M}'\mathbf{M})^{-1}$ exist.

The arbitrariness in a generalized inverse of $\mathbf{A}$ is investigated by means of this partitioning. Thus, on substituting (44) into $\mathbf{AGA} = \mathbf{A},$ we get

$$\mathbf{L}\mathbf{A}_{11}\mathbf{M}\mathbf{G}\mathbf{L}\mathbf{A}_{11}\mathbf{M} = \mathbf{L}\mathbf{A}_{11}\mathbf{M}. \tag{45}$$

Pre-multiplication by $\mathbf{A}_{11}^{-1}(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'$ and post-multiplication by $\mathbf{M}'(\mathbf{M}'\mathbf{M})^{-1}\mathbf{A}_{11}^{-1}$ then gives

$$\mathbf{MGL} = \mathbf{A}_{11}^{-1}. \tag{46}$$

Whatever the generalized inverse is, suppose it is partitioned as

$$\mathbf{G} = \begin{bmatrix} (\mathbf{G}_{11})_{r\times r} & (\mathbf{G}_{12})_{r\times(p-r)} \\ (\mathbf{G}_{21})_{(q-r)\times r} & (\mathbf{G}_{22})_{(q-r)\times(p-r)} \end{bmatrix} \tag{47}$$

of order $q \times p$, conformable for multiplication with $\mathbf{A}$. Then substituting (47) and (44) into (46) gives

$$\mathbf{G}_{11} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}_{21} + \mathbf{G}_{12}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} = \mathbf{A}_{11}^{-1}. \tag{48}$$

This is true whatever the generalized inverse may be. Therefore, on substituting from (48) for $\mathbf{G}_{11}$, we have

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}_{21} - \mathbf{G}_{12}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} \tag{49}$$

as a generalized inverse of $\mathbf{A}$ for any matrices $\mathbf{G}_{12}, \mathbf{G}_{21},$ and $\mathbf{G}_{22}$ of appropriate order. Thus, the arbitrariness of a generalized inverse is characterized.

Example 13 Illustration of the Characterization in (49)

Let $\mathbf{A} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$ and $\mathbf{G} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ -\frac{1}{4} & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix}$. This generalized inverse only satisfies Penrose condition (i). Partition $\mathbf{A}$ so that $\mathbf{A}_{11} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$, $\mathbf{A}_{12} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$,

$\mathbf{A}_{21} = \begin{bmatrix} 2 & 0 \end{bmatrix}$, and $\mathbf{A}_{22} = [2]$. Also $\mathbf{G}_{11} = \begin{bmatrix} \frac{1}{4} & 0 \\ -\frac{1}{4} & \frac{1}{2} \end{bmatrix}$, $\mathbf{G}_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mathbf{G}_{21} = \begin{bmatrix} -\frac{1}{4} & 0 \end{bmatrix}$, and $\mathbf{G}_{22} = \begin{bmatrix} \frac{1}{2} \end{bmatrix}$. Now $\mathbf{A}_{11}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$. Using Formula 49, we can see that

$$\mathbf{G}_{11} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{4} & 0 \end{bmatrix}$$
$$- \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

□

Certain consequences of (49) can be noted.

1. By making $\mathbf{G}_{12}, \mathbf{G}_{21}$, and $\mathbf{G}_{22}$ null, $\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, a form discussed earlier.
2. When $\mathbf{A}$ is symmetric, $\mathbf{G}$ is not necessarily symmetric. Only when $\mathbf{G}_{12} = \mathbf{G}_{21}'$ and $\mathbf{G}_{22}$ is symmetric will $\mathbf{G}$ be symmetric.
3. When $p \geq q$, $\mathbf{G}$ can have full row rank q even if $r < q$. For example, if $\mathbf{G}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}_{22}$, $\mathbf{G}_{21} = \mathbf{0}$ and $\mathbf{G}_{22}$ has full row rank the rank of $\mathbf{G}$ can exceed the rank of $\mathbf{A}$. In particular, this means that singular matrices can have non-singular generalized inverses.

The arbitrariness evident in (49) prompts investigating the relationship of one generalized inverse to another. It is simple. If $\mathbf{G}_1$ is a generalized inverse of $\mathbf{A}$, then so is

$$\mathbf{G} = \mathbf{G}_1\mathbf{A}\mathbf{G}_1 + (\mathbf{I} - \mathbf{G}_1\mathbf{A})\mathbf{X} + \mathbf{Y}(\mathbf{I} - \mathbf{A}\mathbf{G}_1) \tag{50}$$

for any $\mathbf{X}$ and $\mathbf{Y}$. Pre- and post-multiplication of (50) by $\mathbf{A}$ shows that this is so.

The importance of (50) is that it provides a method of generating all generalized inverses of $\mathbf{A}$. They can all be put in the form of (50). To see this, we need only show that for some other generalized inverse $\mathbf{G}_2$ that is different from $\mathbf{G}_1$, there exist values of $\mathbf{X}$ and $\mathbf{Y}$ giving $\mathbf{G} = \mathbf{G}_2$. Putting $\mathbf{X} = \mathbf{G}_2$ and $\mathbf{Y} = \mathbf{G}_1\mathbf{A}\mathbf{G}_2$ achieves this.

The form of $\mathbf{G}$ in (50) is entirely compatible with the partitioned form given in (49). For if we take $\mathbf{G}_1 = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and partition $\mathbf{X}$ and $\mathbf{Y}$ in the same manner as $\mathbf{G}$, then (50) becomes

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{X}_{21} - \mathbf{Y}_{12}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{X}_{22} + \mathbf{Y}_{12} \\ \mathbf{X}_{21} - \mathbf{Y}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{X}_{22} + \mathbf{Y}_{22} \end{bmatrix}. \tag{51}$$

This characterizes the arbitrariness even more specifically than does (49). Thus, for the four sub-matrices of $\mathbf{G}$ shown in (47) we have

Sub-matrix	Source of Arbitrariness
$\mathbf{G}_{11}$	$\mathbf{X}_{21}$ and $\mathbf{Y}_{12}$
$\mathbf{G}_{12}$	X_{22} and Y_{12}
$\mathbf{G}_{21}$	$\mathbf{X}_{21}$ and $\mathbf{Y}_{22}$
$\mathbf{G}_{22}$	$\mathbf{X}_{22}$ and $\mathbf{Y}_{22}$

This means that in the partitioning of

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix}$$

implicit in (50), the first set of rows in the partitioning of $\mathbf{X}$ does not enter into $\mathbf{G}$, and neither does the first set of columns of $\mathbf{Y}$.

It has been shown earlier (Theorem 3) that all solutions to $\mathbf{Ax} = \mathbf{y}$ can be generated from $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z}$, where $\mathbf{z}$ is the infinite set of arbitrary vectors of order q. We now show that all solutions can be generated from $\tilde{\mathbf{x}} = \mathbf{Gy}$ where $\mathbf{G}$ is the infinite set of generalized inverses indicated in (50). First, a Lemma is needed.

Lemma 10 If $\mathbf{z}_{q\times 1}$ is arbitrary and $\mathbf{y}_{p\times 1}$ is known and non-null, there exists an arbitrary matrix $\mathbf{X}$ such that $\mathbf{z} = \mathbf{Xy}$.

Proof. Since $\mathbf{y} \neq \mathbf{0}$ at least one element y_k say, will be non-zero. Writing $\mathbf{z} = \{z_j\}$ and $\mathbf{X} = \{x_{ij}\}$ for $i = 1,\ldots, q$ and $j = 1,\ldots, p$, let $x_{ij} = z_i/y_k$ for $j = k$ and $x_{ij} = 0$ otherwise. Then $\mathbf{Xy} = \mathbf{z}$ and $\mathbf{X}$ is arbitrary. ■

We use this lemma to prove the theorem on generating solutions.

Theorem 11 For all possible generalized inverses $\mathbf{G}$ of $\mathbf{A}$, $\tilde{\mathbf{x}} = \mathbf{Gy}$ generates all solutions to the consistent equations $\mathbf{Ax} = \mathbf{y}$.

Proof. For the generalized inverse $\mathbf{G}_1$, solutions to $\mathbf{Ax} = \mathbf{y}$ are $\tilde{\mathbf{x}} = \mathbf{G}_1\mathbf{y} + (\mathbf{G}_1\mathbf{A} - \mathbf{I})\mathbf{z}$ where $\mathbf{z}$ is arbitrary. Let $\mathbf{z} = -\mathbf{Xy}$ for some arbitrary $\mathbf{X}$**.** Then

$$\begin{aligned}
\tilde{\mathbf{x}} &= \mathbf{G}_1\mathbf{y} - (\mathbf{G}_1\mathbf{A} - \mathbf{I})\mathbf{Xy} \\
&= \mathbf{G}_1\mathbf{y} - \mathbf{G}_1\mathbf{AG}_1\mathbf{y} + \mathbf{G}_1\mathbf{AG}_1\mathbf{y} + (\mathbf{I} - \mathbf{G}_1\mathbf{A})\mathbf{Xy} \\
&= [\mathbf{G}_1\mathbf{AG}_1 + (\mathbf{I} - \mathbf{G}_1\mathbf{A})\mathbf{X} + \mathbf{G}_1(\mathbf{I} - \mathbf{AG}_1)\mathbf{y} \\
&= \mathbf{Gy},
\end{aligned}$$

where $\mathbf{G}$ is exactly the form given in (50) using $\mathbf{G}_1$ for $\mathbf{Y}$. ■

In Theorem 9 (iii), we showed how to represent any generalized inverse of $\mathbf{X'X}$ in terms of the Moore–Penrose inverse. Theorem 12 shows how to do this for a generalized inverse of any matrix $\mathbf{A}$.

Theorem 12 Let $\mathbf{G}$ be any generalized inverse of $\mathbf{A}$. Then

(i) $\mathbf{UU'GS'S} = \mathbf{A}^+$.

(ii) $\mathbf{G} = \mathbf{X}^+ + \mathbf{UC}_1\mathbf{T} + \mathbf{VC}_2\mathbf{S} + \mathbf{VC}_3\mathbf{T} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{-1/2} & \mathbf{C}_1 \\ \mathbf{C}_2 & \mathbf{C}_3 \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{T} \end{bmatrix}$,
where $\mathbf{C}_1 = \mathbf{V'GS'},\ \mathbf{C}_2 = \mathbf{U'GT'}$ and $\mathbf{C}_3 = \mathbf{V'GT'}$.

Proof. (i) Since $\mathbf{AGA} = \mathbf{A}$, we have, using the singular value decomposition of $\mathbf{A}$,

$$\mathbf{S'\Lambda}^{1/2}\mathbf{U'GS'\Lambda}^{1/2}\mathbf{U'} = \mathbf{S'\Lambda}^{1/2}\mathbf{U'}. \tag{52}$$

Pre-multiply (52) by $\mathbf{\Lambda}^{-1/2}\mathbf{S}$ and post-multiply by $\mathbf{U\Lambda}^{-1/2}$. Then we get

$$\mathbf{U'GS'} = \mathbf{\Lambda}^{-1/2}. \tag{53}$$

Pre-multiply (53) by $\mathbf{U}$ and post-multiply by $\mathbf{S}$ to obtain

$$\mathbf{UU'GS'S} = \mathbf{A}^+.$$

(ii) Notice that

$$\begin{aligned} \mathbf{G} &= (\mathbf{UU'} + \mathbf{VV'})\mathbf{G}(\mathbf{S'S} + \mathbf{T'T}) \\ &= \mathbf{UU'GS'S} + \mathbf{VV'GS'S} + \mathbf{UU'GT'T} + \mathbf{VV'GT'T} \\ &= \mathbf{A}^+ + \mathbf{VC}_1\mathbf{S} + \mathbf{UC}_2\mathbf{T} + \mathbf{VC}_3\mathbf{T}. \end{aligned}$$

■

7. OTHER RESULTS

Procedures for inverting partitioned matrices are well-known (e.g., Section 8.7 of Searle (1966), Section 3 of Gruber (2014)). In particular, the inverse of the partitioned full-rank symmetric matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{X'} \\ \mathbf{Z'} \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B'} & \mathbf{D} \end{bmatrix}, \tag{54}$$

say, can for

$$\mathbf{W} = (\mathbf{D} - \mathbf{B'A}^{-1}\mathbf{B})^{-1} = [\mathbf{Z'Z} - \mathbf{Z'X}(\mathbf{X'X})^{-1}\mathbf{X'Z}],$$

be written as

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{W}\mathbf{B}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{W} \\ -\mathbf{W}\mathbf{B}'\mathbf{A}^{-1} & \mathbf{W} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{I} \end{bmatrix} \mathbf{W} \begin{bmatrix} -\mathbf{B}'\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}. \tag{55}$$

The analogy for (55) for generalized inverses, when $\mathbf{M}$ is symmetric but singular, has been derived by Rhode (1965). In defining $\mathbf{A}^-$ and $\mathbf{Q}^-$ as generalized inverses of $\mathbf{A}$ and $\mathbf{Q}$**,** respectively, where $\mathbf{Q} = \mathbf{D} - \mathbf{B}'\mathbf{A}^-\mathbf{B}$, then a generalized inverse of $\mathbf{M}$ is

$$\mathbf{M}^{-} = \begin{bmatrix} \mathbf{A}^{-} + \mathbf{A}^{-}\mathbf{B}\mathbf{Q}^{-}\mathbf{B}'\mathbf{A}^{-} & -\mathbf{A}^{-}\mathbf{B}\mathbf{Q}^{-} \\ -\mathbf{Q}^{-}\mathbf{B}'\mathbf{A}^{-} & \mathbf{Q}^{-} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A}^{-} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{A}^{-}\mathbf{B} \\ \mathbf{I} \end{bmatrix} \mathbf{Q}^{-} \begin{bmatrix} -\mathbf{B}'\mathbf{A}^{-} & \mathbf{I} \end{bmatrix}. \tag{56}$$

It is to be emphasized that the generalized inverses referred to here are just as have been defined throughout, namely satisfying only the first of Penrose's four conditions. (In showing that $\mathbf{M}\mathbf{M}^-\mathbf{M} = \mathbf{M}$, considerable use is made of Theorem 7.)

Example 14 A Generalized Inverse of a Partitioned Matrix
Consider the matrix with the partitioning,

$$\mathbf{M} = \begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 3 \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \mathbf{D}, \mathbf{B} = \mathbf{B}' = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

A generalized inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix},$$
$$\mathbf{Q} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} \end{bmatrix} \text{ and}$$
$$\mathbf{Q}^{-} = \begin{bmatrix} \frac{2}{3} & 0 \\ 0 & 0 \end{bmatrix}.$$

Then,

$$\mathbf{M}^{-} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 0 & 1 \end{bmatrix}$$

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

We could have used different generalized inverses for **A** and **Q**. If we had done so, we would get a valid but different generalized inverse for **M**.

The regular inverse of the product **AB** is $\mathbf{B}^{-1}\mathbf{A}^{-1}$. However, there is no analogous result for generalized inverses. When one matrix is non-singular, **B**, say, Rohde (1964) indicates that $\mathbf{B}^{-1}\mathbf{A}^{-}$ is a generalized inverse of **AB**. Greville (1966) considers the situation for unique generalized inverses $\mathbf{A}^{(\mathrm{p})}$ and $\mathbf{B}^{(\mathrm{p})}$, and gives five separate conditions under which $(\mathbf{AB})^{(\mathrm{p})} = \mathbf{B}^{(\mathrm{p})}\mathbf{A}^{(\mathrm{p})}$. However, one would hope for conditions less complex that those of Greville for generalized inverses $\mathbf{A}^{-}$and $\mathbf{B}^{-}$ satisfying just the first of Penrose's conditions. What can be shown is that $\mathbf{B}^{-}\mathbf{A}^{-}$ is a generalized inverse of **AB** if and only if $\mathbf{A}^{-}\mathbf{ABB}^{-}$ is idempotent. Furthermore, when the product **AB** is itself idempotent, it has **AB**, $\mathbf{AA}^{-}$, $\mathbf{BB}^{-}$, and $\mathbf{B}^{-}\mathbf{BAA}^{-}$ as generalized inverses. Other problems of interest are the generalized inverse of $\mathbf{A}^{k}$ in terms of that of **A**, for integer k, and the generalized inverse of $\mathbf{XX}'$ in terms of that of $\mathbf{X}'\mathbf{X}$.

8. EXERCISES

1 Reduce the matrices

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & -1 \\ 5 & 8 & 0 & 1 \\ 1 & 2 & -2 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & 2 \\ 7 & 8 & 10 & 7 \\ 2 & 1 & 1 & 6 \end{bmatrix}$$

to diagonal form and find a generalized inverse of each.

2 Find generalized inverses of **A** and **B** in Exercise 1 by inverting non-singular minors.

3 For **A** and **B** of Exercise 1, find general solutions to

(a) $\mathbf{AX} = \begin{bmatrix} -1 \\ -13 \\ 11 \end{bmatrix}$

(b) $\mathbf{Bx} = \begin{bmatrix} 14 \\ 23 \\ 32 \\ -5 \end{bmatrix}$

4 Find the Penrose inverse of $\begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 5 \\ 0 & 1 & -1 \\ 1 & 3 & -1 \end{bmatrix}$.

5 Which of the remaining axioms for a Moore–Penrose inverse are satisfied by the generalized inverse in Example 2?

6 **(a)** Using the Algorithm in Section 1b, find generalized inverses of

$$\mathbf{A}_1 = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

derived from inverting the 2×2 minors

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}, \mathbf{M}_2 = \begin{bmatrix} 1 & 15 \\ 1 & 5 \end{bmatrix}, \text{ and } \mathbf{M}_3 = \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}.$$

(b) Using the Algorithm in Section 1b find a generalized inverse of

$$\mathbf{A}_2 = \begin{bmatrix} 2 & 2 & 6 \\ 2 & 3 & 8 \\ 6 & 8 & 22 \end{bmatrix}$$

derived from inverting the minor

$$\mathbf{M} = \begin{bmatrix} 3 & 8 \\ 8 & 22 \end{bmatrix}.$$

7 Let

$$\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

(a) Find the Moore–Penrose inverse of **A.**

(b) Classify the following generalized inverses of **A** as named in Table 1.1 by determining which of the Penrose conditions are satisfied.

(i) $\begin{bmatrix} \frac{3}{4} & \frac{9}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$

(ii) $\begin{bmatrix} \frac{5}{4} & \frac{3}{4} \\ \frac{7}{4} & \frac{1}{4} \end{bmatrix}$

(iii) $\begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ 1 & \frac{1}{2} \end{bmatrix}$

(iv) $\begin{bmatrix} \frac{3}{4} & \frac{5}{4} \\ \frac{5}{4} & \frac{3}{4} \end{bmatrix}$

8 Given $\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

Find

(a) A minimum norm generalized inverse of **X**

(b) A least-square generalized inverse of **X**

(c) The Moore–Penrose inverse of **X**

9 Find a generalized inverse of each of the following matrices.

(a) **PAQ**, when **P** and **Q** are non-singular

(b) **GA**, when **G** is a generalized inverse of **A**

(c) $k\mathbf{A}$, where k is a scalar

(d) **ABA,** when **ABA** is idempotent

(e) **J**, when **J** is square with every element unity

10 What kinds of matrices

(a) are their own generalized inverses?

(b) have transposes as a generalized inverse?

(c) have an identity matrix as a generalized inverse?

(d) have every matrix of order $p \times q$ for a generalized inverse?

(e) have only non-singular generalized inverses?

11 Explain why the equations (a) $\mathbf{Ax} = \mathbf{0}$ and (b) $\mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{y}$ are always consistent.

12 If $\mathbf{z} = (\mathbf{G} - \mathbf{F})\mathbf{y} + (\mathbf{I} - \mathbf{FA})\mathbf{w}$, where **G** and **F** are generalized inverses of **A**, show that the solution $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{GA} - \mathbf{I})\mathbf{z}$ to $\mathbf{Ax} = \mathbf{y}$ reduces to $\tilde{\mathbf{x}} = \mathbf{Fy} + (\mathbf{FA} - \mathbf{I})\mathbf{w}$.

13 If $\mathbf{Ax} = \mathbf{y}$ are consistent equations, and $\mathbf{F}$ and $\mathbf{G}$ are generalized inverses of $\mathbf{A}$, find in simplest form, a solution for $\mathbf{w}$ to the equations

$$(\mathbf{I} - \mathbf{GA})\mathbf{w} = (\mathbf{F} - \mathbf{G})\mathbf{y} + (\mathbf{FA} - \mathbf{I})\mathbf{z}.$$

14 **(a)** If $\mathbf{A}$ has full-column rank**,** show that its generalized inverses are also left inverses satisfying the first three Penrose conditions.

(b) If $\mathbf{A}$ has full-row rank**,** show that its generalized inverses are also right inverses satisfying the first, second, and fourth Penrose conditions.

15 Show that (29) reduces to (27).

16 Give an example of a singular matrix that has a non-singular generalized inverse.

17 Prove that $\mathbf{B}^-\mathbf{A}^-$ is a generalized inverse of $\mathbf{AB}$ if and only if $\mathbf{A}^-\mathbf{ABB}^-$ is idempotent.

18 Show that the rank of a generalized inverse of $\mathbf{A}$ does not necessarily have the same rank as $\mathbf{A}$ and that it is the same if and only if it has a reflexive generalized inverse. See Rhode (1966), also see Ben-Israel and Greville (2003), and Harville (2008).

19 When $\mathbf{PAQ} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ with $\mathbf{P}$ and $\mathbf{Q}$ non-singular show that $\mathbf{G} = \mathbf{Q}\begin{bmatrix} \mathbf{D}^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}\mathbf{P}$ is a generalized inverse of $\mathbf{A}$. Under what conditions does $\mathbf{GAG} = \mathbf{G}$? Use $\mathbf{G}$ to answer Exercise 15.

20 Using $\mathbf{AGA} = \mathbf{A}$

(a) Find a generalized inverse of $\mathbf{AB}$ where $\mathbf{B}$ is orthogonal.

(b) Find a generalized inverse of $\mathbf{LA}$ where $\mathbf{A}$ is non-singular.

21 What is the Penrose inverse of a symmetric idempotent matrix?

22 If $\mathbf{G}$ is a generalized inverse of $\mathbf{A}_{p\times q}$, show that $\mathbf{G} + \mathbf{Z} - \mathbf{GAZAG}$ generates

(a) all generalized inverses of $\mathbf{A}$, and

(b) all solutions to consistent equations $\mathbf{Ax} = \mathbf{y}$ as $\mathbf{Z}$ ranges over all matrices of order $q \times p$.

23 Show that the generalized inverse of $\mathbf{X}$ that was derived in Theorem 12

$$\mathbf{G} = \mathbf{X}^+ + \mathbf{UC}_1\mathbf{T} + \mathbf{VC}_2\mathbf{S} + \mathbf{VC}_3\mathbf{T} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Lambda^{-1/2} & \mathbf{C}_1 \\ \mathbf{C}_2 & \mathbf{C}_3 \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{T} \end{bmatrix}$$

(a) Satisfies Penrose condition (ii)(is reflexive) when $\mathbf{C}_3 = \mathbf{C}_2\mathbf{\Lambda}^{1/2}\mathbf{C}_1$;

(b) Satisfies Penrose condition (iii)(is minimum norm) when $\mathbf{C}_2 = 0$;

(c) Satisfies Penrose Condition (iv)(is a least-square generalized inverse when $\mathbf{C}_1 = 0$.)

24 Show that $\mathbf{M} = (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'$ and $\mathbf{W} = \mathbf{X}'(\mathbf{X}'\mathbf{X})^{+}$ are expressions for the Moore–Penrose inverse of $\mathbf{X}$

(i) by direct verification of the four Penrose conditions.

(ii) using the singular value decomposition.

25 Show that if $\mathbf{N}$ is a non-singular matrix, then $(\mathbf{UNU}')^{+} = \mathbf{UN}^{-1}\mathbf{U}'$.

26 Show that if $\mathbf{P}$ is an orthogonal matrix, $(\mathbf{PAP}')^{+} = \mathbf{PA}^{+}\mathbf{P}'$.

27 Show that

(a) $\mathbf{X}^{+}(\mathbf{X}')^{+} = (\mathbf{X}'\mathbf{X})^{+}$;

(b) $(\mathbf{X}')^{+}\mathbf{X}^{+} = (\mathbf{XX}')^{+}$.

28 Show that the generalized inverses that would be produced by the algorithms in Sections 1a and 1b are reflexive.

29 Show that $\mathbf{K}$ as defined in equation (30) satisfies the four Penrose axioms.

30 Show that if $\mathbf{X}^{-}$ satisfies Penrose's condition (iv) then $\mathbf{b} = \mathbf{X}^{-}\mathbf{y}$ is a solution to $\mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{y}$. [*Hint:* use Exercise 22 or Theorem 12.]

31 Show that $\mathbf{M}^{-}$ of (56) is a generalized inverse of $\mathbf{M}$ in (54).

32 If $\mathbf{P}_{m\times q}$ and $\mathbf{D}_{m\times m}$ have rank m show that $\mathbf{D}^{-1} = \mathbf{P}(\mathbf{P}'\mathbf{DP})^{-}\mathbf{P}'$.

33 Show by direct multiplication that

$$\mathbf{M}^{-} = \begin{bmatrix} 0 & 0 \\ 0 & (\mathbf{Z}'\mathbf{Z})^{-} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -(\mathbf{Z}'\mathbf{Z})^{-}\mathbf{Z}'\mathbf{X} \end{bmatrix} \mathbf{Q}^{-} \begin{bmatrix} \mathbf{I} & -\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-} \end{bmatrix},$$

where $\mathbf{Q} = \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-}\mathbf{Z}'\mathbf{X}$ is a generalized inverse of

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}.$$