Generalized Inverses and Generalized Connections with Statistics

Tim Zitzer

April 12, 2007

Generalized Inverses and Generalized Connections with Statistics

Consider an arbitrary system of linear equations with a coefficient matrix $A \in \mathbb{M}_{n,n}$, vector of constants $b \in \mathbb{C}^n$, and solution vector $x \in \mathbb{C}^n$:

$Ax = b$

If matrix A is nonsingular and thus invertible, then we can employ the techniques of matrix inversion and multiplication to find the solution vector x. In other words, the unique solution $x = A^{-1}b$ exists. However, when A is rectangular, dimension $m \times n$ or singular, a simple representation of a solution in terms of A is more difficult. There may be none, one, or an infinite number of solutions depending on whether $b \in C(A)$ and whether $n - rank(A) > 0$. We would like to be able to find a matrix (or matrices) G, such that solutions of $Ax = b$ are of the form Gb. Thus, through the work of mathematicians such as Penrose, the generalized inverse matrix was born. In broad terms, a generalized inverse matrix of A is some matrix G such that Gb is a solution to $Ax = b$.

Definitions and Notation

In 1920, Moore published the first work on generalized inverses. After his definition was more or less forgotten due to cumbersome notation, an algebraic form of Moore's definition was given by Penrose in 1955, having explored the theoretical properties in depth. We give the Penrose definition as it will be referred to frequently.

Definition *Penrose Generalized Inverse*

For any matrix $A \in \mathbb{M}_{m,n}$ there exists a matrix $G \in \mathbb{M}_{m,n}$ satisfying the conditions

$$\begin{array}{ll} (i) & AGA = A \\ (ii) & GAG = G \\ (iii) & (AG)^* = AG \\ (iv) & (GA)^* = GA \end{array}$$

where M^* is the conjugate transpose of the square matrix M.

There are other types of generalized inverses which are defined by satisfying only some of the conditions given above, but we will focus on the generalized inverse. As with any new concept or idea, the first important fact to be established is the existence and uniqueness of a generalized inverse matrix. We begin with existence and uniqueness will follow.

Theorem 0.0.1 Existence of Generalized Inverse *For any matrix A, there is a matrix A^g, which has its rows and columns in the row space and column space of A^T and also satisfies equations (i)-(iv) from the definition above.*

The proof of this existence theorem is lengthy and is not included here. For the interested reader, the proof can be found in the article by T. N. E. Greville *The Pseudoinverse of a Rectangular or Singular Matrix and its Application to the Solution of Systems of Linear Equations*, where pseudoinverse is synonymous with generalized inverse. Next we address the uniqueness of a generalized inverse, so we can refer to a matrix A^g as **the** generalized inverse.

Theorem 0.0.2 Uniqueness of Generalized Inverse
Given a matrix $A \in \mathbb{M}_{m,n}$, there exists a matrix G, such that G provides a unique solution to equations (i) - (iv) given above.

Proof To establish the uniqueness of A^g, we show that if G is any matrix satisfying the defining equations for A^g, then $GAA^* = A^*$ and $G^* = AB_1$ for some matrix B_1. First we can combine condition (i) and (iii) from the Penrose definition to show that $GAA^* = A^*$. The conjugate transpose of (ii) using (iv) gives $G^* = G^*A^*G^* = A(GG^*) = AB_1$ for $B_1 = GG^*$. Likewise, G satisfies the same two equations for another matrix B_2. Using $GAA^* = A^*$, we have

$(G - A^g)AA^* = 0$ or $(G - A^g)A = 0$

Using $G^* = AB_1$, $(G-A^g) = (B_1^* - B_2^*)A^*$ which implies $(G-A^g)C = 0$, where the matrix C has columns orthogonal to the columns of A. The equations $(G-A^g)AA^* = 0$ or $(G-A^g)A = 0$ and $(G - A^g)C = 0$ together imply that $G - A^g = 0$, which establishes the uniqueness of A^g. □

It is interesting to note that the Penrose definition includes the case $A = \mathcal{O}$, since the generalized inverse of the zero matrix, $\mathcal{O}_{m,n}$ is equal to its transpose $\mathcal{O}_{n,m}$. We can also relate the generalized inverse back to a regular inverse matrix. Essentially the regular inverse matrix is a special case of a generalized inverse matrix. We bring to light the connection in the following theorem.

Theorem 0.0.3 Regular Inverse and Generalized Inverse
If A is a nonsingular matrix of size n, then A^{-1} is the generalized inverse matrix guaranteed in the Penrose definition by the existence and uniqueness theorems.

Proof We check the four conditions listed in the Penrose definition, letting $G = A^{-1}$.
(i) $AA^{-1}A = I_nA = A$ as desired.
(ii) $A^{-1}AA^{-1} = I_nA^{-1} = A^{-1}$ as desired.
(iii) $(AA^{-1})^* = (I_n)^* = I_n$ since the identity matrix is real and composed of real numbers.
(iv) $(A^{-1}A)^* = (I_n)^* = I_n$ by the same reasoning as above.
Thus the inverse A^{-1} is actually the generalized inverse of a nonsingular matrix A.

Methods of Calculation

Most of the writing on the computational aspects of generalized inverses is highly theoretical, only dealing with alternative algebraic structures or interesting properties of the various types of generalized inverses. This is likely due to the fact that the usefulness of generalized inverses, in statistical areas especially, is theoretical rather than practical [Pringle]. With that, the calculation of A^g from a matrix A can be difficult. We present an example of the generalized inverse for a simple 3×2 matrix and let the reader verify conditions (i)-(iv) from the Penrose definition.

Example 3×2 *Generalized Inverse*
Given the 3×2 matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 2 \end{bmatrix}$, without any justification, the generalized inverse is given by $A^g = \begin{bmatrix} \frac{1}{30} & \frac{1}{15} & \frac{1}{30} \\ \frac{1}{15} & \frac{2}{15} & \frac{1}{15} \end{bmatrix}$.

We now present a method which will enable us to begin to calculate generalized inverses of small matrices like the one given in the previous example. This method should help give a feeling of what constructing A^g entails, using the functional definition included in S.L. Campbell's *Generalized Inverses of Linear Transformations*. Though not presented here, it is instructive to follow for the linear algebra involved, while the interested reader can research the motivation behind this method.

Example *Campbell's Method*

$$\text{Let } A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & 2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Then the column space of A^*, denoted $C(A^*)$ is spanned by the basis

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Using results from Linear Algebra, a basis of $C(A^*)$ can be shown to be

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Now,

$$A \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 2 \\ 2 \end{bmatrix} \text{ and } A \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 1 \end{bmatrix}. \text{ Thus } A^g \begin{bmatrix} 3 \\ 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ and } A^g \begin{bmatrix} 3 \\ 4 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

since $A^g Ax = x$ if $x \in C(A^*) = C(A^g)$ as shown in Campbell's functional definition.
The null space of A^* is

$$\left\langle \left\{ \begin{bmatrix} -1 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ \frac{1}{2} \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$

Then $A^g \begin{bmatrix} -1 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} = A^g \begin{bmatrix} -1 \\ \frac{1}{2} \\ 0 \\ 1 \end{bmatrix} = \vec{0} \in \mathbb{C}^3$ since $A^g x = 0$ for $x \in \mathcal{N}(A^*)$ also as shown in Campbell's functional definition.
Combining all of this gives

$$A^g \begin{bmatrix} 3 & 3 & -1 & -1 \\ 2 & 4 & \frac{1}{2} & \frac{1}{2} \\ 2 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \text{ or } A^g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 3 & -1 & -1 \\ 2 & 4 & \frac{1}{2} & \frac{1}{2} \\ 2 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix}^{-1}$$

So $A^g = \begin{bmatrix} \frac{1}{7} & -\frac{5}{21} & \frac{11}{42} & \frac{11}{42} \\ 0 & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} \\ \frac{1}{7} & \frac{2}{21} & \frac{2}{21} & \frac{2}{21} \end{bmatrix}$

Note the indicated inverse can always be computed since its columns form a basis for $C(A) \oplus \mathcal{N}(A^*)$ and hence are a linearly independent set of four vectors in $\mathbb{C}^4$. This direct sum result is proven in Campbell's *Generalized Inverses* and we encourage the interested reader to study it carefully, but we will not include it here. Below is a formal statement of the methods employed above.

Theorem 0.0.4 Campbell's Method Computation
Let $A_{m\times n} \in \mathbb{C}^{m\times n}$ have rank r. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a basis for $C(A^*)$ and $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n-r}\}$ is a basis for $\mathcal{N}(A^*)$, then

$$A^g = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_r & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \ldots & A\mathbf{v}_r & \mathbf{w}_1 & \mathbf{w}_2 & \ldots & \mathbf{w}_{n-r} \end{bmatrix}^{-1}$$

where the entries in the matrices above represent columns.

Proof By using the properties of matrix-vector multiplication across columns and the property $A^gAx = 0$ if $x \in \mathcal{N}(A)$ and $A^gAx = x$ if $x \in C(A^*) = C(A^g)$, we have

$$\begin{aligned} A^g \begin{bmatrix} A\mathbf{v}_1 & \ldots & A\mathbf{v}_r & \mathbf{w}_1 & \ldots & \mathbf{w}_{n-r} \end{bmatrix} &= \begin{bmatrix} A^gA\mathbf{v}_1 & \ldots & A^gA\mathbf{v}_r & A^g\mathbf{w}_1 & \ldots & A^g\mathbf{w}_{n-r} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_r & 0 & \ldots & 0 \end{bmatrix} \end{aligned}$$

Again the entries in the matrices above actually represent column vectors.
Furthermore, $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ must be a basis for $C(A)$ since the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is linearly independent and has size equal to the number of columns of A^* with leading 1's. Thus, since A^* has n rows, and the basis for $C(A^*)$ has r vectors, then multiplying each of the vectors in the basis on the left by A will yield a set of r $(m \times 1)$ vectors. Since matrix A was defined to have rank r, then we have the correct dimension and linear independence so the set $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ is a basis for $C(A)$. Similarly, since the set of all $x \in C^n$ that yield the zero vector when the inner product is computed with any element of $C(A)$ is equal to $\mathcal{N}(A^*)$, it follows that the matrix

$$\begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \ldots & A\mathbf{v}_r & \mathbf{w}_1 & \mathbf{w}_2 & \ldots & \mathbf{w}_{n-r} \end{bmatrix}$$

must be nonsingular. (See Campbell's Proposition 0.2.1 for the proof that the set of vectors $u \in \mathbb{C}^n$ such that $\langle u, v\rangle = 0$ for every $v \in C(A)$ equals the null space of A). The desired result is now immediate. $\square$

One other method to compute the generalized inverse uses an idea suggested by A.S. Householder, in a method developed by Thomas Greville, which can be found in *Generalized Inverses Theory and Applications* by Ben-Israel and Greville. We encourage the interested reader to study the proof carefully, but here we only give a proposition. Recall that the rank is equal to the number of pivot columns or the number of nonzero rows, and other equivalent properties. Using the rank we can decompose a rectangular matrix into two pieces in order to compute the generalized inverse.

Proposition 0.0.5 Rank Factorization
If $A \in \mathbb{C}^{m\times n}$, then there exists $B \in \mathbb{C}^{m\times r}, C \in \mathbb{C}^{r\times n}$, such that $A = BC$ and $r = rank(A) = rank(B) = rank(C)$.

Theorem 0.0.6 Rank Decomposition
*Let r be the rank of a $m \times n$ matrix A and $A = BC$, where B and C are the matrices given by the **Rank Factorization Proposition**. Then, $A^g = C^g B^g$ where $B^g = (B^*B)^{-1}B^*$ is the generalized inverse for matrix B and $C^g = C^*(CC^*)^{-1}$ is the generalized inverse for matrix C.*

Before we prove the **Rank Decomposition Theorem**, we give some lemmas regarding conjugation and inversion. These are given without proof as they are trivial, but for the interested reader, look to Robert Beezer's *A First Course in Linear Algebra* text.

Lemma 0.0.7 Conjugate of Conjugation
For any matrix A, $(A^*)^* = A$.

Lemma 0.0.8 Conjugate of Inverse
For a nonsingular matrix A, $(A^{-1})^* = (A^*)^{-1}$.

Lemma 0.0.9 Conjugate of Product
For any $m \times n$ matrix A and any $n \times p$ matrix B, $(AB)^* = B^*A^*$.

These lemmas play a large part in the following proof.

Proof We now check the four conditions listed in the Penrose definition.

(i) $B(B^*B)^{-1}B^*B$
$= B(B^*B)^{-1}(B^*B)$
$= BI_k = B$

(ii) $(B^*B)^{-1}B^*B(B^*B)^{-1}B^*$
$= (B^*B)^{-1}(B^*B)(B^*B)^{-1}B^*$
$= I_k(B^*B){-}1B^*$
$= (B^*B)^{-1}B^* = B^g$

(iii) $(B(B^*B)^{-1}B^*)^*$
$= ((B((B^*B)^{-1}))(B^*))^*$ using **Conjugation of Product and Inverse lemmas**
$= B(B((B^*B)^{-1}))^*$
$= (B((B^*B)^{-1})B^*) = BB^g$

(iv) $((B^*B)^{-1}B^*B)^*$
$= (((B^*B)^{-1})^*(B^*B))^* =$ using **Conjugation of Product and Inverse lemmas**
$= (((B^*B)^{-1})(B^*B))$
$= ((B^*B)^{-1}B^*B) = B^gB$

The proof of checking conditions (i)-(iv) for the Penrose definition of the generalized inverse for matrix C^g is very similar to the proof for B^g so we will leave this to the reader to carefully prove.
After verifying the conditions for C^g, B^g and C^g satisfy the four conditions listed in the Penrose definition. Now, using the factorization of A, we can express A^g as

$$A^g = (BC)^g = C^gB^g = C^*(CC^*)^{-1}(B^*B)^{-1}B^* \qquad \square$$

This result is similar to regular inverses of nonsingular matrices–if we had $A = BC$, where A, B, C, are all nonsingular matrices, then taking the inverse of both sides reverses the order of B and C to give $A^{-1} = C^{-1}B^{-1}$. However, we must be very careful since $(BC)^g = C^g B^g$ does not hold in general, unless B is of full column rank and C is of full row rank, since in that case we have

$(BC)^g = C^*(CC^*)^{-1}(B^*B)^{-1}B^* = C^g B^g$ (as shown above)

and we can apply the **Rank Factorization Proposition**.

We can extend this method with two special cases, both can be thought of as A being halfway nonsingular. Here are the corollaries to the **Rank Decomposition Theorem**.

Corollary 0.0.10 Full Row Rank Decomposition
In our initial decomposition of $A = BC$, if A has full row rank, in other words $r = m$, then B can be chosen to be the identity matrix I_m, and C is now of dimension $m \times n$. The formula of the decomposition then becomes $A = I_m C$, and the formula for A^g reduces to

$$A^g = A^*(AA^*)^{-1}$$

Corollary 0.0.11 Full Column Rank Decomposition
In our initial decomposition of A into $A = BC$, if A has full column rank, so that $r = n$, then C can be chosen to be the identity matrix I_n, and B is now of dimension $m \times n$. The decomposition then becomes $A = BI_n$, and the formula for A^g reduces to

$$A^g = (A^*A)^{-1}A^*$$

The proofs of these are straightforward and can be verified simply by plugging into the Rank Decomposition theorem. We now carefully study an example of the decomposition technique detailed above.

Example *Rank Decomposition*
Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \end{bmatrix}$

The rank of A is $r = 1$, since A row-reduces to $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ and has only one leading 1. We decompose A into $A = BC$ where $B \in \mathbb{M}_{2,1}$ and $C \in \mathbb{M}_{1,3}$. We can extract the column vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ to show with little thought that $A = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$.

Then $B^*B = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \end{bmatrix}$, and $CC^* = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \end{bmatrix}$.

Thus $A^g = C^*(CC^*)^{-1}(B^*B)^{-1}B^* = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} \frac{1}{6} \end{bmatrix} \begin{bmatrix} \frac{1}{5} \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \frac{1}{30} \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{30} & \frac{1}{15} \\ \frac{1}{30} & \frac{1}{15} \\ \frac{1}{15} & \frac{2}{15} \end{bmatrix}$

The method of computing A^g described in the **Rank Decomposition Theorem** may be executed by reducing A with elementary row operations. This formalized process is implemented as an efficient algorithm for calculating generalized inverses. But first, we give a new definition.

Definition *Row-Echelon Form*
A matrix $E \in \mathbb{M}_{m,n}$ which has rank r is said to be in **row echelon form** if E is of the form

$$E = \left[\begin{array}{c} C_{r\times n} \\ \hline 0_{(m-r)\times n} \end{array}\right]$$

where the elements c_{ij} of $C(= C_{r\times n})$ satisfy the following conditions,

(i) $c_{ij} = 0$ when $i > j$
(ii) The first non-zero entry in each row of C is 1
(iii) If $c_{ij} = 1$ is the first non-zero entry of the ith row, then the jth column of C is the unit vector e_i whose only non-zero entry is in the ith position.

Example *REF*
The matrix

$$E = \begin{bmatrix} 1 & 2 & 0 & 3 & 5 & 0 & 1 \\ 0 & 0 & 1 & -2 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in row-echelon form as it satisfies the conditions stated in the definition above.

Below we state some facts about the row-echelon form, the proofs of which are somewhat trivial in relation to generalized inverses, so we encourage the interested reader on his own time to find the proofs in Ben Noble's *Applied Linear Algebra*.

Properties 0.0.12 Row Echelon Form
For $A \in \mathbb{C}_{m,n}$ such that rank$(A) = r$:

(E1) A can always be row-reduced to row-echelon form by elementary row operations (i.e. there always exists a non-singular matrix $P \in \mathbb{M}_{m,m}$ such that $PA = E_A$ where E_A is in row-echelon form).

(E2) For a given A, the row-echelon form E_A obtained by row-reducing A is unique.

(E3) If E_A is the row-echelon form for A and the unit vectors in E_A appear in columns $i_1, i_2, \ldots,$ and i_r, then the corresponding columns of A are a basis for $C(A)$. This particular basis is called the set of distinguished columns or pivot columns of A. The remaining columns are called the undistinguished columns of A. For example, the matrix E has pivot columns for the first, third, and sixth columns.

*(E4) If E_A is the row-echelon from the **Row-Echelon Form Definition** for A, then $\mathcal{N}(A) = \mathcal{N}(E_A) = \mathcal{N}(C)$.*

*(E5) If the partitioned matrix in **Definition Row-Echelon Form** is the row echelon form for A, and if $B \in \mathbb{C}^{m\times r}$ is the matrix made up of the pivot columns of A (in the same order as they are in A), then $A = BC$ where C is obtained from the partitioned row*

echelon form as in the definition above. This is known as a full rank factorization such as was detailed in the ***Rank Factorization Proposition*** *above.*

We can proceed to define an algorithm, adapted from S.L. Campbell's *Generalized Inverses of Linear Transformations*, to obtain the full rank factorization and generalized inverse for any $A \in \mathbb{M}_{m,n}$. Here we go.

Algorithm *Rank Factorization*

(I) Reduce A to row echelon form E_A.

(II) Select the pivot columns of A and place them as the columns in a matrix B in the same order as they appear in A.

(III) Select the non-zero rows from E_A and place them as rows in a matrix C in the same order as they appear in E_A.

(IV) Compute $(CC^*)^{-1}$and$(B^*B)^{-1}$.

(V) Compute A^g as $A^g = C^*(CC^*)^{-1}(B^*B)^{-1}B^*$.

Example *REF Rank Factorization*

Using the algorithm above, we find A^g where

$$A = \begin{bmatrix} 1 & 2 & 1 & 4 & 1 \\ 2 & 4 & 0 & 6 & 8 \\ 1 & 2 & 0 & 3 & 4 \\ 3 & 6 & 0 & 9 & 12 \end{bmatrix}$$

(I) Using elementary row operations we reduce A to its row echelon form

$$E_A = \begin{bmatrix} 1 & 2 & 0 & 3 & 4 \\ 0 & 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(II) The first and third columns are pivot columns. Thus

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

(III) The matrix C is made up of the non-zero rows of E_A so that

$$C = \begin{bmatrix} 1 & 2 & 0 & 3 & 4 \\ 0 & 0 & 1 & 1 & -3 \end{bmatrix}$$

(IV) Now $CC^* = \begin{bmatrix} 30 & -9 \\ -9 & 11 \end{bmatrix}$ and $B^*B = \begin{bmatrix} 15 & 1 \\ 1 & 1 \end{bmatrix}$. Calculating $(CC^*)^{-1}$ and $(B^*B)^{-1}$ we get $(CC^*)^{-1} = \frac{1}{249}\begin{bmatrix} 11 & 9 \\ 9 & 30 \end{bmatrix}$ and $(B^*B)^{-1} = \frac{1}{14}\begin{bmatrix} 1 & -1 \\ -1 & 15 \end{bmatrix}$.

(V) Substituting the results of steps (II), (III), and (IV) into the formula for A^g gives

$$A^g = C^*(CC^*)^{-1}(B^*B)^{-1}B^* = \frac{1}{3486}\begin{bmatrix} 126 & 4 & 2 & 6 \\ 252 & 8 & 4 & 12 \\ 420 & -42 & -21 & -63 \\ 798 & -30 & -15 & -45 \\ -756 & 142 & 71 & 213 \end{bmatrix}$$

Applications and Connections

We finally turn to the connection with exactly what kind of solutions generalized inverses give. First let us recall some notation and a simple lemma. We denote the Euclidean norm of a vector w, or the inner product $\langle w, w\rangle$, as $||w||$.

Lemma 0.0.13 Norm of Sum
If $u, v \in \mathbb{C}^p$ *and* $\langle u, v\rangle = 0$, *then* $||u+v||^2 = ||u||^2 + ||v||^2$.

Proof Supposing the hypotheses given in the Lemma above, then $||u+v||^2 = \langle u+v, u+v\rangle = \langle u, u\rangle + \langle v, u\rangle + \langle u, v\rangle + \langle v, v\rangle = ||u||^2 + ||v||^2$.

Let us revisit the problem of finding solutions $\mathbf{x}$ to $Ax = b$, $A \in \mathbb{M}_{m,n}, b \in \mathbb{C}^m$. If this system of equations is overdetermined, we can still look for a vector u that minimizes $Au - b$ or $||Au - b||^2$. The following definition and theorem is adapted from Gene Golub's *Matrix Computations* and Campbell's work.

Definition *Least Squares Solution*
Suppose that $A \in \mathbb{M}_{m,n}$ and $b \in \mathbb{C}^m$. Then a vector $u \in \mathbb{C}^n$ is called a **least squares solution** to $Ax = b$ if $||Au - b|| \leq ||Av - b||$ for all $v \in \mathbb{C}^n$. A vector u is called a **minimal least squares solution** to $Ax = b$ if u is a least squares solution to $Ax = b$ and $||u|| < ||w||$ for all other least squares solutions w.

The name least squares is derived from the definition of the Euclidean norm as the square root of the sum of squares. If $b \in C(A)$, then the concepts of solution and least squares solution obviously coincide as we will see in the next theorems.

Theorem 0.0.14 Least Squares Solution
Suppose A is an $m \times n$, usually with $m > n$ in the case of minimizing residuals for a given plot, x is an unknown n-dimensional parameter vector and b is a known m-dimensional measurement vector. If we want to minimize the Euclidean norm of the residual, $||Ax - b||$, then a solution vector x which minimizes the residual is given by $x = (A^*A)^{-1}A^*b = A^g b$ for a matrix A of full column rank.

Proof *Adapted from Bapat's Linear Algebra & Linear Models*
Using the fact that $||v||^2$ is equal to v^*v, we rewrite $||Ax - b||$ as $(Ax - b)^*(Ax - b) = (Ax)^*(Ax) - b^*Ax - (Ax)^*b + b^*b$. The two middle terms are equal, as can be shown easily, again we encourage the reader to do so, and the minimum of this can be found by taking the derivative of $A^*A(x)^2 - 2A^*bx$ with respect to x, setting the derivative equal to zero, and solving for the minimizing vector x. Thus, the equation is $2A^*Ax - 2A^*b = 0$, which has

a minimum at $A^*Ax = A^*b$. This is now a system of linear equations, but by hypothesis, we presumed that A is of full column rank, and since A^*A is a square matrix, we also know that A^*A is invertible. Thus, a solution of the system of linear equations is given by $x = (A^*A)^{-1}A^*b = A^g b$, where $A^g = (A^*A)^{-1}A^*$ by the **Full Column Rank Decomposition Corollary**. $\square$

Since A^g is unique, the minimal least squares solution is a unique solution, but we will be more precise in the next theorem.

Theorem 0.0.15 Least Squares and Generalized Inverse Uniqueness
Suppose that $A \in \mathbb{M}_{m,n}$ and $b \in \mathbb{C}^m$. Then $A^g b$ is ***the*** *minimal least squares solution to $Ax = b$.*

Proof Notice that $||Ax - b||^2 = ||Ax - AA^g b \oplus -(I - AA^g)b||^2 = ||Ax - AA^g b||^2$.
Thus x will be a least squares solution if and only if x is a solution of the consistent system $Ax = AA^g b$. But solutions of $Ax = AA^g b$ are of the form $x = A^g(AA^g b) \oplus (I - A^g A)b = A^g b \oplus (I - A^g A)b$. Since $||x||^2 = ||A^g b||^2$ we see that there is exactly one minimal least squares solution $x = A^g b$. $\square$

If minimality is not as important, then the next theorem can be very useful.

Theorem 0.0.16 Forms of Least Squares Solutions
Suppose that $A \in \mathbb{M}_{m,n}$ and $b \in \mathbb{C}^m$. Then the following statements are equivalent
(i) u is a least squares solution of $Ax = b$
(ii) u is a solution of $Ax = AA^g b$
(iii) u is a solution of $A^*Ax = A^*b$
(iv) u is of the form $A^g b + h$ where $h \in \mathcal{N}(A)$

Proof We know from the proof of Theorem Least Squares and Generalized Inverses that (i), (ii), and (iv) are equivalent. If (i) holds, then multiplying $Au = b$ on the left by A^* gives (iii). On the other hand, multiplying $A^*Au = A^*b$ on the left by A^{*g} gives $Au = AA^g b$. Thus (iii) implies (ii). $\square$

We present one last example here to show the usefulness of generalized inverses in statistics. The linear least squares method detailed above can be used to find a function that best fits a given set of data.

Example *Linear Least Squares Function*
Consider the Cartesian points $(0, 3), (2, 3), (4, 4), (-1, 2)$. We seek a solution of the form $\alpha x + \beta = y$, that is $\begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = y$

We can then form the matrix A using the x-coordinates and a column of 1's as $A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \\ -1 & 1 \end{bmatrix}$

and the vector b using the y-coordinates $b = \begin{bmatrix} 3 \\ 3 \\ 4 \\ 2 \end{bmatrix}$.

Now we can find $A^*, A^*b, A^*A, \text{and}(A^*A)^{-1}$

$$A^* = \begin{bmatrix} 0 & 2 & 4 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad A^*b = \begin{bmatrix} 20 \\ 12 \end{bmatrix} \quad A^*A = \begin{bmatrix} 21 & 5 \\ 5 & 4 \end{bmatrix} \quad (A^*A)^{-1} = \tfrac{1}{59} \begin{bmatrix} 4 & -5 \\ -5 & 21 \end{bmatrix}$$

Using the equation $A^*Ax = A^*b$ from the **Least Squares Solution** proof, we can simply substitute in to obtain

$$A^*A \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = A^*b = \begin{bmatrix} 21 & 5 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 20 \\ 12 \end{bmatrix}$$

and $\frac{1}{59} \begin{bmatrix} 4 & -5 \\ -5 & 21 \end{bmatrix} \begin{bmatrix} 20 \\ 12 \end{bmatrix} = \begin{bmatrix} \frac{20}{59} \\ \frac{152}{59} \end{bmatrix}$ and the line of best fit is $\frac{20}{59}x + \frac{152}{59} = y$.

With that we have shown the usefulness of the generalized inverse in finding the least squares function for a given set of data. The generalized inverse is a powerful tool and has even more connections with statistics than we can possibly bring to light in the confines of this paper. Please consult the sources listed for further reference and additions to the theorems and techniques provided here.

Sources

Bapat, R.B. *Linear Algebra & Linear Models.* New Delhi: Springer-Verlag, 2000.

Beezer, Rob. *A First Course in Linear Algebra.* University of Puget Sound: GNU Free Documentation License, 2007.

Ben-Israel, Adi & Greville, Thomas N.E. *Generalized Inverses Theory and Applications.* New York: Springer-Verlag, 2003, 1974.

Bouillion, Thomas L. *Generalized Inverse Matrices.* New York: Wiley-Interscience, 1971.

Campbell, S.L. & Meyer, C.D. Jr. *Generalized Inverses of Linear Transformations.* London: Pitman Publishing Ltd., 1979.

Golub, Gene H. & Van Loan, Charles F. *Matrix Computations.* Baltimore: Johns Hopkins University Press, 1983.

Ortega, James M. *Matrix Theory: A 2nd Course.* New York: Plenum Press, 1987.

Pringle, R.M. & Rayner, A.A. *Generalized Inverse Matrices with Applications to Statistics.* London: Charles Griffin & Co. Ltd., 1971.

Rao, C. Rahakrishna & Mitra, Sujit Kumar. *Generalized Inverse of Matrices and its Applications.* New York: John Wiley & Sons, Inc. 1971.