What Is the Generalized Inverse of a Matrix?

Jeff Gill, jgill@ucdavis.edu

[Edited out of Gill and King: "What to do When Your Hessian is Not Invertible: Alternatives to Model Respecification in Nonlinear Estimation."]

1 The Generalized Inverse

This research note briefly reviews the mathematical and statistical considerations behind the generalized inverse. The literature on the theory and application of the generalized inverse is vast and spans several fields. Here we summarize some of the fundamental principles. See Harville (1997) for further details.

1.1 The Many Generalized Inverse Matrices

Any matrix, $\mathbf{A}$, can be decomposed as

$$\underset{(p\times q)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{L}}\ \underset{(p\times q)}{\mathbb{D}}\ \underset{(q\times q)}{\mathbf{U}} \qquad \text{where,} \qquad \mathbb{D} = \begin{bmatrix} \mathbf{D}_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}, \tag{1.1}$$

and both $\mathbf{L}$ (lower triangular) and $\mathbf{U}$ (upper triangular) are non-singular (even given a singular $\mathbf{A}$). The diagonal matrix $\mathbf{D}_{r\times r}$ has dimension and rank r corresponding to the rank of $\mathbf{A}$. When $\mathbf{A}$ is non-negative definite and symmetric, then the diagonals of $\mathbf{D}_{r\times r}$ are the eigenvalues of $\mathbf{A}$. If $\mathbf{A}$ is non-singular, positive definite, and symmetric, as in the case of a proper invertible Hessian, then $\mathbf{D}_{r\times r} = \mathbb{D}$ (i.e. $r = q$) and $\mathbf{A} = \mathbf{L}\mathbb{D}\mathbf{L}'$. The matrices $\mathbf{L}$, $\mathbb{D}$, and $\mathbf{U}$ are all non-unique unless $\mathbf{A}$ is nonsingular.

By rearranging 1.1 we can diagonalize any matrix as

$$\mathbb{D} = \mathbf{L}^{-1}\mathbf{A}\mathbf{U}^{-1} = \begin{bmatrix} \mathbf{D}_{\mathbf{r}\times\mathbf{r}} & 0 \\ 0 & 0 \end{bmatrix}. \tag{1.2}$$

Now define a new matrix, $\mathbb{D}^-$ created by taking the inverses of the non-zero (diagonal) elements of $\mathbb{D}$:

$$\mathbb{D}^- = \begin{bmatrix} \mathbf{D}^-_{\mathbf{r}\times\mathbf{r}} & 0 \\ 0 & 0 \end{bmatrix}. \tag{1.3}$$

If $\mathbb{D}\mathbb{D}^- = \mathbf{I}_{\mathbf{q}\times\mathbf{q}}$ then we could say that $\mathbb{D}^-$ is *the* inverse of $\mathbb{D}$. However, this is not true:

$$\mathbb{D}\mathbb{D}^- = \begin{bmatrix} \mathbf{D}_{\mathbf{r}\times\mathbf{r}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{D}^-_{\mathbf{r}\times\mathbf{r}} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 \\ 0 & 0 \end{bmatrix}$$

Instead, we notice that:

$$\mathbb{D}\mathbb{D}^-\mathbb{D} = \begin{bmatrix} \mathbf{1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{\mathbf{r}\times\mathbf{r}} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{\mathbf{r}\times\mathbf{r}} & 0 \\ 0 & 0 \end{bmatrix} = \mathbb{D}.$$

So $\mathbb{D}^-$ is *a* generalized inverse of $\mathbb{D}$ because of the extra structure required. Note that this is *a* generalized inverse not *the* generalized inverse since the matrices on the right side of (1.1) are non-unique. By rearranging (1.1) and using (1.3) we can define a new $q \times p$ matrix: $\mathbf{G} = \mathbf{U}^{-1}\mathbb{D}^-\mathbf{L}^{-1}$. The importance of the *generalized inverse* matrix $\mathbf{G}$ is revealed in the following theorem.[1]

[1] The "generalized inverse" is also sometimes referred to as the "conditional inverse", "pseudo inverse", and "g-inverse".

Theorem 1.1 *(Moore 1920)* $\mathbf{G}$ *is a generalized inverse of* $\mathbf{A}$ *since* $\mathbf{AGA} = \mathbf{A}$

The new matrix $\mathbf{G}$ necessarily has rank r since the product rule states that the result has rank less than or equal to the minimum of the rank of the factors, and $\mathbf{AGA} = \mathbf{A}$ requires that $\mathbf{A}$ must have rank less than or equal to the lowest rank of itself or $\mathbf{G}$.

Although $\mathbf{G}$ has infinitely many definitions that satisfy Theorem 1.1, any one of them will do for our purposes. For example, in linear regression, the fitted values — defined as $X\mathbf{G}X'y$, with $\mathbf{G}$ as the generalized inverse of $X'X$, X as a matrix of explanatory variables and y as the outcome variable — are invariant to the definition of $\mathbf{G}$. In addition, we only use our pseudo-variance as a first approximation to the surface of the true posterior, and we will improve it in our importance resampling stage.[2]

1.2 The Unique Moore-Penrose Generalized Inverse Matrix

Moore (1920) and (unaware of Moore's work) Penrose (1955) reduced the infinity of generalized inverses to one unique solution by imposing four reasonable algebraic constraints, all met by the standard inverse. If

1. general condition: $\mathbf{AGA} = \mathbf{A}$,
2. reflexive condition: $\mathbf{GAG} = \mathbf{G}$,
3. normalized condition: $(\mathbf{AG})' = \mathbf{GA}$, and
4. reverse normalized condition: $(\mathbf{GA})' = \mathbf{AG}$

then this $\mathbf{G}$ matrix is unique. The proof is lengthy, and we refer the interested reader to Penrose (1955). There is a vast literature on generalized inverses that meet some subset of the Moore-Penrose condition. A matrix that satisfies the first two conditions is called a "reflexive" or "weak" generalized inverse and is order dependent. A matrix that satisfies the first three conditions is called a "normalized" generalized inverse. A matrix that satisfies the first and fourth conditions is called a "minimum norm" generalized inverse.

Because the properties of the Moore-Penrose generalized inverse are intuitively desirable, and because of the invariance of important statistical results to the choice of generalized inverse, we follow standard statistical practice by using this form from now on. The implementations of the generalized inverse in Gauss and Splus are both the Moore-Penrose version.

The Moore-Penrose generalized inverse is also easy to calculate using QR factorization. QR factorization takes the input matrix, $\mathbf{A}$, and factors it into the product of an orthogonal matrix, $\mathbf{Q}$, and a matrix, $\mathbf{R}$, which has a triangular leading square matrix ($\mathbf{r}$) followed by rows of zeros corresponding to the difference in rank and dimension in $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{r} \\ \mathbf{0} \end{bmatrix}.$$

This factorization is implemented in virtually every professional level statistical package. The Moore-Penrose generalized inverse is produced by:

$$\mathbf{G} = \begin{bmatrix} \mathbf{r}^{-1}\mathbf{0} \end{bmatrix} \mathbf{Q}'$$

where $\mathbf{0}$ is the transpose of the zeros portion of the $\mathbf{R}$ matrix required for conformability.

[2]Note in addition that $\mathbf{AG}$ is always idempotent ($\mathbf{GAGA} = \mathbf{G}(\mathbf{AGA}) = \mathbf{GA}$), and rank($\mathbf{AG}$) = rank($\mathbf{A}$). These results hold whether $\mathbf{A}$ is singular or not.

1.3 Numerical Examples

As a means of motivating a simple numerical example of how the generalized inverse works, we develop a brief application to the linear model where the $\mathbf{X}'\mathbf{X}$ matrix is noninvertible because $\mathbf{X}$ is singular. In this context, the generalized inverse provides a solution to the normal equations (Campbell and Meyer 1979, p.94), and both the fitted values of $\mathbf{Y}$ and the residual error variance are invariant to the choice of $\mathbf{G}$ (Searle 1971, p.169-71). We use the Moore Penrose generalized inverse.

Let

$$\mathbf{X} = \begin{bmatrix} 5 & 2 & 5 \\ 2 & 1 & 2 \\ 3 & 2 & 3 \\ 2.95 & 1 & 3 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 9 \\ 11 \\ -5 \\ -2 \end{bmatrix}$$

(Our omission of the constant term makes the numerical calculations cleaner but is not material to our points.) Applying the least squares model to these data ($\mathbf{X}$ is of full rank) yields the coefficient vector

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y = (222.22, -11.89, -215.22)',$$

fitted values,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}} = (11.22, 2.11, -2.78, -2.00)'.$$

and variance matrix

$$\Sigma = \begin{bmatrix} 57283.95 & -1580.25 & -56395.06 \\ -1580.25 & 187.65 & 1491.36 \\ -56395.06 & 1491.36 & 55550.62 \end{bmatrix}.$$

What we call the *standardized correlation matrix*, a correlation with standard deviations on the diagonal, is then

$$\mathbf{C}_s = \begin{bmatrix} 239.34 & -0.48 & -0.99 \\ -0.48 & 13.69 & 0.46 \\ -0.99 & 0.46 & 235.69 \end{bmatrix}.$$

Now suppose we have a matrix of explanatory affects that is identical to $\mathbf{X}$ except that we have changed the bottom left number from 2.95 to 2.99

$$\mathbf{X_2} = \begin{bmatrix} 5 & 2 & 5 \\ 2 & 1 & 2 \\ 3 & 2 & 3 \\ 2.99 & 1 & 3 \end{bmatrix}$$

Using the same $\mathbf{Y}$ outcome vector and applying the same least squares calculation now gives

$$\hat{\mathbf{b}}_\mathbf{2} = (1111.11, -11.89, -1104.11)',$$

and

$$\hat{\mathbf{Y}} = (11.22, 2.11, -2.78, -2.00)'.$$

However, the variance-covariance matrix reacts sharply to the movement towards singularity as seen in the standardized correlation matrix:

$$\mathbf{C}_s = \begin{bmatrix} 1758.79 & -0.48 & -0.99 \\ -0.48 & 20.12 & 0.48 \\ -0.99 & 0.48 & 1753.35 \end{bmatrix}$$

Indeed, if $\mathbf{X_3} = 2.999$, then $\mathbf{X'X}$ is singular (with regard to precision in Gauss and Splus) and we must use the generalized inverse. This produces

$$\tilde{\mathbf{b}}_{\mathbf{3}} = \mathbf{GX'Y} = (1.774866, -5.762093, 1.778596)',$$

and

$$\hat{\mathbf{Y}} = \mathbf{XGX'Y} = (6.2431253, 1.3448314, -0.8637996, 4.8965190)'.$$

The resulting pseudo-variance matrix (calculated now from $\mathbf{G}\sigma^2$) produces larger standard deviations for the first and third explanatory variable, reflecting greater uncertainty, again displayed as a standardized correlation matrix:

$$\mathbf{C}_s = \begin{bmatrix} 16328.7257311 & -0.4822391 & -0.9999999 \\ -0.4822391 & 18.6815417 & 0.4818444 \\ -0.9999999 & 0.4818444 & 16323.6599450 \end{bmatrix}$$

References

Campbell, S. L. and C. D. Meyer, Jr. 1979. *Generalized Inverses of Linear Transformations.* New York: Dover Publications.

Gill, Phillip E., G. H. Golub, Walter Murray, and M. A. Sanders. 1974. "Methods for Modifying Matrix Factorizations." *Mathematics of Computation* 28, 505-35.

Gill, Phillip E. and Walter Murray. 1974. "Newton-Type Methods for Unconstrained and Linearly Constrained Optimization." *Mathematical Programming* 7, 311-50.

Gill, Phillip E., Walter Murray, and M. H. Wright. 1981. *Practical Optimization.* London: Academic Press.

Harville, David A. *Matrix Algebra From a Statistician's Perspective.* New York: Springer.

Matthews, A. and D. Davies. 1971. "A Comparison of Modified Newton Methods for Unconstrained Optimization." *Computer Journal* 14, 213-94.

Moore, E. H. 1920. "On the Reciprocal of the General Algebraic Matrix. (Abstract)" *Bulletin of the American Mathematical Society* 26, 394-5.

Penrose, R. A. 1955. "A Generalized Inverse for Matrices." *Proceedings of the Cambridge Philosophical Society* 51, 406-13.

Searle, S. R. 1971. *Linear Models.* New York: Wiley & Sons.