Generalized Inverses:
How to Invert a Non-Invertible Matrix

S. Sawyer — September 7, 2006 rev August 6, 2008

1. Introduction and Definition. Let A be a general $m \times n$ matrix. Then a natural question is when we can solve

$$Ax = y \qquad \text{for } x \in R^m, \qquad \text{given } y \in R^n \tag{1.1}$$

If A is a square matrix ($m = n$) and A has an inverse, then (1.1) holds if and only if $x = A^{-1}y$. This gives a complete answer if A is invertible. However, A may be $m \times n$ with $m \neq n$, or A may be a square matrix that is not invertible.

If A is not invertible, then equation (1.1) may have no solutions (that is, y may be not be in the range of A), and if there are solutions, then there may be many different solutions.

For example, assume $A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$. Then $A\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, so that A is not invertible. It would be useful to have a characterization of those $y \in R^2$ for which it is possible to find a solution of $Ax = y$, and, if $Ax = y$ is a solution, to find all possible solutions. It is easy to answer these questions directly for a 2×2 matrix, but not if A were 8×3 or 10×30.

A solution of these questions can be found in general from the notion of a *generalized inverse* of a matrix:

Definition. If A is an $m \times n$ matrix, then G is a *generalized inverse* of A if G is an $n \times m$ matrix with

$$AGA = A \tag{1.2}$$

If A has an inverse in the usual sense, that is if A is $n \times n$ and has a two-sided inverse A^{-1}, then

$$A^{-1}(AGA)A^{-1} = (A^{-1}A)\,G\,(AA^{-1}) = G$$

while by (1.2)

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

Thus, if A^{-1} exists in the usual sense, then $G = A^{-1}$. This justifies the term *generalized* inverse. We will see later that any $m \times n$ matrix A has at least one generalized inverse G. However, unless A is $n \times n$ and A is invertible, there are many different generalized inverses G, so that G is not unique. (Generalized inverses are unique is you impose more conditions on G; see Section 3 below.)

One consequence of (1.2) is that AGAG=AG and GAGA=GA. In general, a square matrix P that satisfies $P^2 = P$ is called a *projection matrix*. Thus both AG and GA are projection matrices. Since A is $m \times n$ and G is $n \times m$, AG is an $m \times m$ projection matrix and GA is $n \times n$.

In general if P is a projection matrix, then $P = P^2$ implies $Py = P(Py)$ and $Pz = z$ for for all $z = Py$ in the range of P. That is, if P is $n \times n$, P moves any $x \in R^n$ into $V = \{Px : x \in R^n\}$ (the range of P) and then keeps it at the same place.

If $x \in R^n$, then $y = Px$ and $z = x - Px = (I - P)x$ satisfies $x = y + z$, $Py = y$, and $Pz = P(x - Px) = Px - P^2x = 0$. Since then $Px = P(y + z) = y$, we can say that P projects R^n *onto* its range V *along* the space $W = \{x : Px = 0\}$.

The two projections AG and GA both appear in the next result, which shows how generalized inverses can be used to solve matrix equations.

Theorem 1.1. Let A by an $m \times n$ matrix and assume that G is a generalized inverse of A (that is, $AGA = A$). Then, for any fixed $y \in R^m$,

(i) the equation

$$Ax = y, \qquad x \in R^n \tag{1.3}$$

has a solution $x \in R^n$ if and only if $AGy = y$ (that is, if and only if y is in the range of the projection AG).

(ii) If $Ax = y$ has any solutions, then x is a solution of $Ax = y$ if and only if

$$x = Gy + (I - GA)z \qquad \text{for some } z \in R^n \tag{1.4}$$

Remark. If we want a particular solution of $Ax = y$ for y in the range of A, we can take $x = Gy$.

Proof of Theorem 1.1. All of the parts of the theorem are easy to prove, but some involve somewhat unintuitive manipulations of matrices.

Proof of part (i): If y is in the range of the projection AG, that is if $(AG)y = y$, then $A(Gy) = y$ and $x = Gy$ is a solution of $Ax = y$. Conversely, if $Ax = y$, then $GAx = Gy$ and $AGAx = AGy = (AG)y$, while $AGA = A$ implies $AGAx = Ax = y$. Thus $(AG)y = y$. Thus, if $Ax = y$ has any solutions for a given $y \in R^m$, then $x = Gy$ is a particular solution.

Proof of part (ii): This has two parts: First, if $AGy = y$, then all of the vectors in (1.4) are solutions of $Ax = y$. Second, that (1.4) contains all possible solutions of $Ax = y$.

If $AGy = y$ and $x = Gy + (I - GA)z$, then $Ax = AGy + A(I - GA)z = y + (A - AGA)z = y$, so that any $x \in R^n$ that satisfies (1.4) with $AGy = y$ is a solution of $Ax = y$.

Conversely, if $Ax = y$, let $z = x$. Then the right-hand side of (1.4) is $Gy+(I-GA)x = Gy+x-G(Ax) = Gy+x-Gy = x$, so that any solution x of $Ax = y$ is given by (1.4) with $z = x$.

Example. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$ as before. Set $G = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Then

$$AGA = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = A$$

so that G is a generalized inverse of A. The two projections appearing in Theorem 1.1 are

$$AG = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix} \quad \text{and} \quad GA = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$$

In this case

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + 2y \\ 3x + 6y \end{pmatrix} = (x + 2y) \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

Thus $Ax = y$ has a solution x only if $y = c \begin{pmatrix} 1 \\ 3 \end{pmatrix}$. On the other hand,

$$AG \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 3x \end{pmatrix} = x \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

so that the range of the projection AG is exactly the set of vectors $\{ c \begin{pmatrix} 1 \\ 3 \end{pmatrix} \}$. The theorem then says that if $y = c \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, then the set of solutions of $Ax = y$ is exactly

$$\begin{aligned} x &= Gc \begin{pmatrix} 1 \\ 3 \end{pmatrix} + (I - GA) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= c \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= c \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = c \begin{pmatrix} 1 \\ 0 \end{pmatrix} + z_2 \begin{pmatrix} -2 \\ 1 \end{pmatrix} \end{aligned} \tag{1.5}$$

It is easy to check that $Ax = y = c \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ for all x in (1.5), and, with some extra work, that these are all solutions.

2. The $ABCD$-Theorem and Generalized Inverses of Arbitrary Matrices. Let A be an arbitrary $m \times n$ matrix; that is, with n columns and m rows. Then we can write

$$A \;=\; \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \;=\; \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix} \;=\; \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ \dots \\ w_m \end{pmatrix}$$

where v_i are the columns of A and w_j are the rows.

In general, the *column rank* of A (call it r_c) is the dimension of the vector space in R^m that is spanned by the columns $\{v_i\}$ of A, and the *row rank* of A (call it r_r) is the dimension of the vector space in R^n that is spanned by the rows $\{w_j\}$ of A. That is, r_c is the largest number of linearly-independent columns v_i in R^m, and r_r is the the largest number of linearly-independent rows w_j in R^n. Then $r_c \le m$, since the largest number of linearly independent vectors in R^m is m, and $r_c \le n$ since there are only n columns to begin with. Thus $r_c \le \min\{m, n\}$. By the same arguments, $r_r \le \min\{m, n\}$.

It can be shown that $r_c = r_r$ for any $m \times n$ matrix, so that the row rank and the column rank of an arbitrary matrix A are the same. The common value $r = r_c = r_r \le \min\{m, n\}$ is called the *rank* of A.

Let A be an $m \times n$ matrix with $\operatorname{rank}(A) = r \le \min\{m, n\}$ as above. Then, one can show that, after a suitable rearrangement of rows and columns, A can be written in partitioned form as

$$A \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{2.1}$$

where a is $r \times r$ and invertible, b is $r \times (n-r)$, c is $(m-r) \times r$, and d is $(m-r) \times (n-r)$. In fact, we can prove the following representation theorem for general matrices:

Theorem 2.1. Let A is an $m \times n$ matrix with rank $r = \operatorname{rank}(A)$. Then the rows and columns can be permuted so that it can be written in the partitioned form (2.1) where a is $r \times r$ and invertible. In that case $d = ca^{-1}b$, so that

$$A \;=\; \begin{pmatrix} a & b \\ c & ca^{-1}b \end{pmatrix} \tag{2.2}$$

(Note that a, b, c, d in (2.1) and (2.2) are *matrices*, not numbers. Some of the entries b, c, d in (2.1) may be empty, in which case they do not appear, for example if $m = n$ and A is invertible.)

Remarks. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2×2 matrix of numbers with $a > 0$ but $r = \mathrm{rank}(A) = 1$, then $\det(A) = ad - bc = 0$. This implies $d = bc/a$. We cannot write bc/a for matrices, but (2.2) with $d = ba^{-1}c$ is the appropriate generalization for matrices. The matrix $d = ca^{-1}b$ is always defined and is $(m-r) \times (n-r)$, since c is $(m-r) \times r$, a^{-1} is $r \times r$, and b is $r \times (n-r)$.

Example. Let $A = xy'$ be the outer product of vectors $x \in R^m$ and $y \in R^n$, so that A is $m \times n$. Assume $x_1 \neq 0$ and $y_1 \neq 0$. Then $\mathrm{rank}(A) = 1$ since every row of A is a multiple of y and every column of A is a multiple of x. In this case, we can write

$$A \;=\; xy' \;=\; \begin{pmatrix} x_1 y_1 & x_1 \begin{pmatrix} y_2 & \dots & y_n \end{pmatrix} \\ y_1 \begin{pmatrix} x_2 \\ \dots \\ x_m \end{pmatrix} & \begin{pmatrix} x_2 y_2 & \dots & x_2 y_n \\ \dots & \dots & \dots \\ x_m y_2 & \dots & x_m y_n \end{pmatrix} \end{pmatrix}$$

This is in the form (2.2) where $b = x_1 \begin{pmatrix} y_2 & \dots & y_n \end{pmatrix}$ is a $1 \times (n-1)$ row vector, $c = y_1 \begin{pmatrix} x_2 & \dots & x_m \end{pmatrix}'$ is an $(m-1) \times 1$ column vector, and

$$d = ba^{-1}c = y_1 \begin{pmatrix} x_2 \\ \dots \\ x_m \end{pmatrix} \frac{1}{x_1 y_1} x_1 \begin{pmatrix} y_2 & \dots & y_n \end{pmatrix} = \begin{pmatrix} x_2 \\ \dots \\ x_m \end{pmatrix} \begin{pmatrix} y_2 & \dots & y_n \end{pmatrix}$$

is the outer product of an $(m-1)$-dimensional vector and an $(n-1)$-dimensional vector.

Remark. Note that (2.2) can also be written

$$A \;=\; \begin{pmatrix} I_r \\ ca^{-1} \end{pmatrix} \begin{pmatrix} a & b \end{pmatrix} \;=\; \begin{pmatrix} a \\ c \end{pmatrix} a^{-1} \begin{pmatrix} a & b \end{pmatrix} \;=\; \begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} I_r & a^{-1}b \end{pmatrix}$$

This can be viewed as a generalization of the representation $A = uv'$ for an outer product of two vectors u, v.

Proof of Theorem 2.1. If the first r rows of A are linearly independent and $\mathrm{rank}(A) = \mathrm{rank}(a) = r$ in (2.1), then the last $m-r$ rows of A are linear combinations of the first r rows. This means that we can write the last $m-r$ rows of A as

$$(c \quad d)_i = \sum_{j=1}^{r} T_{ij}\, (a \quad b)_j \qquad \text{for } 1 \leq i \leq m-r$$

where T_{ij} $(1 \le i \le m-r,\ 1 \le j \le r)$ are numbers. In terms of matrices,

$$(c \quad d) \;=\; T\,(a \quad b) \;=\; (Ta \quad Tb) \tag{2.3}$$

where T is $(m-r)\times r$.

The relation (2.3) implies $c = Ta$ and hence $T = ca^{-1}$. This implies $Tb = ca^{-1}b = d$ in (2.3), which completes the proof of Theorem 2.1.

Theorem 2.2. Let

$$A \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;=\; \begin{pmatrix} a & b \\ c & ca^{-1}b \end{pmatrix} \tag{2.4}$$

be an $m \times n$ matrix with $r = \operatorname{rank}(A)$ where a is $r \times r$ and invertible, as in Theorem 2.1. Let

$$G \;=\; \begin{pmatrix} a^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{2.5}$$

where the "0"s in (2.5) represent matrices of zeroes of dimension sufficient to make G an $n \times m$ matrix. Then G is a generalized inverse of A.

Proof. By (2.4) and (2.5)

$$\begin{aligned} &\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a^{-1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} I_r & 0 \\ ca^{-1} & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} \;=\; \begin{pmatrix} a & b \\ c & ca^{-1}b \end{pmatrix} \end{aligned}$$

where I_r is the $r \times r$ unit matrix. This implies $AGA = A$ since $d = ca^{-1}b$ by (2.4), so that G is a generalized inverse of A.

The two projections in this case are

$$AG = \begin{pmatrix} I_r & 0 \\ ca^{-1} & 0 \end{pmatrix} \quad \text{and} \quad GA = \begin{pmatrix} I_r & a^{-1}b \\ 0 & 0 \end{pmatrix}$$

Theorem 1.1 then says that $Ax = y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ can be solved for $y_1 \in R^r$, $y_2 \in R^{m-r}$ if and only if

$$AGy \;=\; \begin{pmatrix} I_r & 0 \\ ca^{-1} & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \;=\; \begin{pmatrix} y_1 \\ ca^{-1}y_1 \end{pmatrix} \;=\; \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

That is, if and only if $y_2 = ca^{-1}y_1$. In that case, the general solution of $Ax = y$ for $x \in R^n$ is

$$\begin{aligned} x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Gy + (I_m - GA)z \\ &= \begin{pmatrix} a^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 & -a^{-1}b \\ 0 & I_{m-r} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= \begin{pmatrix} a^{-1}y_1 \\ 0 \end{pmatrix} + \begin{pmatrix} -a^{-1}bz_2 \\ z_2 \end{pmatrix} \end{aligned}$$

for arbitrary $z_2 \in R^{m-r}$.

Remark. This shows that any $m \times n$ matrix A has at least one generalized inverse G of the form (2.5). Since often many different linearly-independent sets of r rows can be permuted to the upper r rows and many different linearly-independent sets of r columns can be permuted into the first r column positions, a matrix A with $\text{rank}(A) = r < n$ can have many different generalized inverses of this form.

3. The Penrose Inverse. In general, an $m \times n$ matrix A has many different generalized inverses unless $m = n$ and A is invertible. It is possible, however, to add conditions to the definition of a generalized inverse so that there is always a unique generalized inverse under the additional conditions.

Definition. G is called a *Penrose inverse* of the $m \times n$ matrix A if G is an $n \times m$ matrix that satisfies the four conditions

(i) $AGA = A$
(ii) $GAG = G$
(iii) $AG = (AG)'$ is an orthogonal projection in R^m
(iv) $GA = (GA)'$ is an orthogonal projection in R^n

Condition (ii) says that A is a generalized inverse of G, in addition to G being a generalized inverse of A.

The fact than an arbitrary $m \times n$ matrix A has a unique $n \times m$ Penrose inverse follows from the Singular Value Decomposition theorem in matrix algebra. Some generalized inverses that are natural to use in practice are Penrose inverses and some are not. The next section gives an example of a Penrose inverse.

4. "Fitted Values" in Statistics. Let X be an $n \times r$ matrix with $r < n$ and $\text{rank}(X) = r$. Then $X'X$ is invertible. If "observed values" $Y \in R^n$ can be "exactly fit" by the parameters $\beta \in R^r$, then

$$Y = X\beta, \qquad Y \in R^n, \;\; \beta \in R^r \tag{4.1}$$

The matrix X cannot be invertible, since $r < n$. However, suppose that we want a general procedure to choose an arbitrary β in terms of Y, in the hopes that later we can find a justification for this procedure other than it gives a definite answer. In that case, we can consider a generalized inverse of X. Specifically, G will be a generalized inverse of X if G is $r \times n$ and

$$XGX = X$$

Since $X'X$ is invertible, an obvious choice is

$$G = (X'X)^{-1}X' \tag{4.2}$$

since then $XGX = X(X'X)^{-1}X'X = X$. The two projections XG and GX are

$$GX = (X'X)^{-1}X'X = I_r \quad \text{and} \quad XG = X(X'X)^{-1}X' = H \tag{4.3}$$

Note that both projections are symmetric: That is, $I_r = (I_r)'$ and $H = H'$. In addition

$$GXG = ((X'X)^{-1}X')X((X'X)^{-1}X') = (X'X)^{-1}X' = G$$

That is, G is the unique Penrose inverse of the $n \times r$ matrix X.

Theorem 1.1 now says that $Y = X\beta$ can be solved exactly if and only if $(XG)Y = HY = Y$; that is, if and only if Y is in the range of the $n \times n$ projection H. Moreover, if $HY = Y$, then every solution of $X\beta = Y$ is of the form

$$\begin{aligned} \beta = GY &= (X'X)^{-1}X'Y + (I_r - GX)z, \quad z \in R^r \\ &= (X'X)^{-1}X'Y \end{aligned} \tag{4.4}$$

since $GX = I_r$ by (4.3). In other words, if $Y = X\beta$ for some vector β, then the only solution β of $X\beta = Y$ for a given Y is given by (4.4).

Indeed, it follows directly from (4.3) that X must be one-one: That is, if $X\beta_1 = X\beta_2$, then $GX\beta_1 = GX\beta_2 = \beta_1 = \beta_2$.

There is a better motivation for the solution $\beta = GY$ for G in (4.2) than arbitrariness (or orneriness). Suppose that we view Y as $X\beta$ that are observed with errors. That is, as

$$Y = X\beta + e \tag{4.5}$$

where $e = \{e_i\}$ are independent errors. Then we can consider the value of β that minimizes the sum of errors

$$\min_{\beta} \sum_{i=1}^{n} \bigl(Y_i - (X\beta)_i\bigr)^2 \;=\; \sum_{i=1}^{n} \bigl(Y_i - (X\widehat{\beta})_i\bigr)^2 \tag{4.6}$$

If we set

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \bigl(Y_i - (X\beta)_i\bigr)^2 \;=\; \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \Bigl(Y_i - \sum_{a=1}^{r} X_{ia}\beta_a\Bigr)^2$$
$$=\; -2\sum_{i=1}^{n} \Bigl(Y_i - \sum_{a=1}^{r} X_{ia}\beta_a\Bigr) X_{ij} \;=\; 0$$

This implies

$$\sum_{a=1}^{r} \sum_{i=1}^{n} X_{ij} X_{ia} \beta_a \;=\; \sum_{i=1}^{n} X_{ij} Y_i, \qquad 1 \le j \le r$$

which can be written in a more compact form as

$$(X'X)\beta = X'Y \tag{4.7}$$

Since we are assuming that $X'X$ is invertible, (4.7) implies

$$\widehat{\beta} = (X'X)^{-1}X'Y = GY$$

for G in (4.2). That is, the *least-squares* solution of (4.6) for β is given by $\beta = \widehat{\beta} = GY$, where G is the Penrose inverse of the $n \times r$ matrix X.