Let $A$ be a given $n \times m$ matrix and $y$ a given $n \times 1$ vector. Consider the linear equation

$$Ab = y.$$  \hspace{1cm} (4.1)

Our task is to find such a vector $b \in \mathbb{R}^m$ that $Ab = y$ is satisfied. It is well known that (4.1) has

(a) no solution at all, or
(b) a unique solution, or
(c) infinite number of solutions.

Moreover, it is easy to conclude the following:

- The equation $Ab = y$ has a solution, i.e., it is consistent, if and only if $y$ belongs to $\mathcal{C}(A)$.

If a solution exists, it is unique if and only if the columns of $A$ are linearly independent, i.e., $A$ has full column rank. Then $(A'A)^{-1}$ exists and premultiplying (4.1) by $(A'A)^{-1}A'$ yields

$$b_0 = (A'A)^{-1}A'y := G_1y.$$ \hspace{1cm} (4.2)

It is obvious that the unique solution to (4.1) can be expressed also as

$$b_0 = (A'NA)^{-1}A'Ny := G_2y,$$ \hspace{1cm} (4.3)

where $N$ an arbitrary $n \times n$ matrix satisfying condition rank$(A'NA) = m$. Note that it is possible that $G_1y = G_2y$ but $G_1 \neq G_2$. If $A$ is a nonsingular square matrix, then of course $b_0 = A^{-1}y$.

We can now try to find such an $m \times n$ matrix $G$, which would behave like $A^{-1}$ as much as possible; for example, we might wish that $G$ has such a property that if (4.1) is consistent, then $Gy$ is one solution. Then $G$ would work like $A^{-1}$ while solving a linear equation; corresponding to $A^{-1}y$ we have $Gy$ as a solution to $Ab = y$. 

_S. Puntanen et al., Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty, DOI 10.1007/978-3-642-10473-2_5, © Springer-Verlag Berlin Heidelberg 2011_
We might emphasize (and repeat) that $\mathbf{A}\mathbf{b} = \mathbf{y}$ does not necessarily have a solution in which case $\mathbf{G}$ does not help in finding a nonexistent solution.

We can define the generalized inverse $\mathbf{G}_{m \times n}$ of matrix $\mathbf{A}_{n \times m}$ in three equivalent ways, see Rao & Mitra (1971b, pp. 20–21):

**Theorem 4.** The matrix $\mathbf{G}_{m \times n}$ is a generalized inverse of $\mathbf{A}_{n \times m}$ if any of the following equivalent conditions holds:

(a) The vector $\mathbf{G}\mathbf{y}$ is a solution to $\mathbf{A}\mathbf{b} = \mathbf{y}$ always when this equation is consistent, i.e., always when $\mathbf{y} \in \mathbb{C}(\mathbf{A})$.

(b) (b1) $\mathbf{G}\mathbf{A}$ is idempotent and $\text{rank}(\mathbf{G}\mathbf{A}) = \text{rank}(\mathbf{A})$, or equivalently

   (b2) $\mathbf{A}\mathbf{G}$ is idempotent and $\text{rank}(\mathbf{A}\mathbf{G}) = \text{rank}(\mathbf{A})$.

(c) $\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}$.

**Proof.** We first show that that (c) and (b2) are equivalent, i.e,

$$\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A} \iff \mathbf{A}\mathbf{G} \text{ is idempotent and } \text{rank}(\mathbf{A}\mathbf{G}) = \text{rank}(\mathbf{A}).$$

Assume first that $\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}$. Then of course $\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}\mathbf{G}$, so that $\mathbf{A}\mathbf{G}$ is idempotent. Because

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{G}\mathbf{A}) \leq \text{rank}(\mathbf{A}\mathbf{G}) \leq \text{rank}(\mathbf{A}),$$

we necessarily have $\text{rank}(\mathbf{A}\mathbf{G}) = \text{rank}(\mathbf{A})$. To go the other way round, assume that $\mathbf{A}\mathbf{G}$ is idempotent and $\text{rank}(\mathbf{A}\mathbf{G}) = \text{rank}(\mathbf{A})$. Then, in view of the rank cancellation rule, Theorem 6 (p. 145), we can cancel the underlined terms from the following equation:

$$\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}\mathbf{G}.$$ (4.6)

Hence (c) implies (b2) and “(c) $\iff$ (b2)” is confirmed. The proof of “(c) $\iff$ (b1)” is quite analogous.

To prove “(a) $\iff$ (c)”, let the vector $\mathbf{G}\mathbf{y}$ be a solution to (4.1) always when (4.1) is consistent, i.e.,

$$\mathbf{A}\mathbf{G}\mathbf{y} = \mathbf{y} \text{ for all } \mathbf{y} \in \mathbb{C}(\mathbf{A}),$$

i.e.,

$$\mathbf{A}\mathbf{G}\mathbf{A}\mathbf{u} = \mathbf{A}\mathbf{u} \text{ for all } \mathbf{u} \in \mathbb{R}^m,$$

which obviously is equivalent to (c). $\square$

A generalized inverse is not necessarily unique—it always exists but it is unique if and only if $\mathbf{A}^{-1}$ exists. If $\mathbf{A}^{-1}$ exists, then pre- and postmultiplying (mp1) by $\mathbf{A}^{-1}$ we get $\mathbf{G} = \mathbf{A}^{-1}$.

The set of all generalized inverses of $\mathbf{A}$ is denoted as $\{\mathbf{A}^{-}\}$:

$$\{\mathbf{A}^{-}\} = \{ \mathbf{G} : \mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A} \}.$$ (4.9)