Jared Andreatta
MATH 531 Homework 3 Spring 2025

About G-inverses

In linear models theory a G-Inverse is a matrix that is used to give an equation for the projection using X when X does not have full rank. From Christensen's book (4^{th} edition, page 428).

Definition B.36. A generalized inverse of any symmetric matrix A is any matrix G such that $AGA = A$. The notation A^- is used to indicate a generalized inverse of A.

One can always construct a G-inverse based on the singular value decomposition (see Theorem B.38.) although there are other choices. For our purposes we are interested in the generalized inverse of X^TX and has the property that $X(X^TX)^-X^T$ is a projection matrix onto the subspace spanned by the columns of X. Note that if X has full rank then the G-inverse is the usual inverse and we obtain the usual projection matrix formula. The G-inverse is useful to derive covariances formulas for estimable parameters but computation is done differently.

1. Suppose $\lambda^T\beta$ is an estimable function, i.e., λ^T is in the row space of X (or the column space of X^T). For convenience let $\mathcal{R}(A)$ denote the row vector space of A and $\mathcal{C}(A)$ denote the column space of A. Then show that

   (a) $\mathcal{R}(\boldsymbol{X}) = \mathcal{R}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) = \mathcal{C}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$.

   Proof. To show that $\mathcal{R}(X) = \mathcal{R}(X^TX)$, we first argue that $\mathcal{R}(X^TX) \subseteq \mathcal{R}(X)$. It is clear that the rows of X^TX can be written as linear combinations of X, since it is essentially a product of X with itself. This implies that vectors in the row space of X^TX is also in X. Therefore, $\mathcal{R}(X^TX) \subseteq \mathcal{R}(X)$.

   Since we've shown that $\mathcal{R}(X^TX) \subseteq \mathcal{R}(X)$, we can say that $\dim \mathcal{R}(X^TX) \leq \dim \mathcal{R}(X)$. However, we know that by the rank preservation property, the matrix X has the same rank as its corresponding Gram matrix, that is, $\mathrm{rank}(X) = \mathrm{rank}(X^TX)$, which means $\dim \mathcal{R}(X^TX) = \dim \mathcal{R}(X)$. This implies that the space spanned by $\mathcal{R}(X^TX)$ must also span $\mathcal{R}(X)$. Therefore, we conclude that $\mathcal{R}(X) = \mathcal{R}(X^TX)$.

   Next, we already know that the matrix X^TX is symmetric. Clearly, it would follow that the rows of X^TX would span the same space as the columns of X^TX. Therefore, we can say that $\mathcal{R}(X^TX) = \mathcal{C}(X^TX)$.

   Hence, we have shown that $\mathcal{R}(X) = \mathcal{R}(X^TX) = \mathcal{C}(X^TX)$. $\square$

   (b) $\lambda^{\mathrm{T}} \in \mathcal{R}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$ if and only if $\lambda^{\mathrm{T}}G\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \lambda^{\mathrm{T}}$, where G is any generalized inverse of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$.

   Proof. Assume $\lambda^T \in \mathcal{R}(X^TX)$. We know that λ^T can be expressed as a linear combination of the row vectors of the matrix X^TX, that is, there is some vector $r^T \in \mathcal{R}(X^TX)$ such that $\lambda^T = r^TX^TX$. We can right-multiply both sides of this equation by GX^TX to find

$$\lambda^TGX^TX = r^TX^TXGX^TX$$

G is a generalized inverse denoted as $(X^TX)^-$, so this simplifies to

$$\lambda^T G X^T X = r^T X^T X$$

We defined $\lambda^T = r^T X^T X$, so the expression simplifies to

$$\lambda^T G X^T X = \lambda^T$$

Hence, $\lambda^T \in \mathcal{R}(X^TX) \implies \lambda^T G X^T X = \lambda^T$ Now, assume that $\lambda^T G X^T X = \lambda^T$. If we examine the LHS of the equation, we notice that the term $\lambda^T G$ is a $1 \times n$ row vector. We call this term r^T. We can substitute back into the equation to find that $r^T X^T X = \lambda^T$. We know that $\mathcal{R}(G) = \mathcal{R}(X^TX)$ since G is just the generalized inverse of X^TX. Clearly, λ^T can be written as a linear combination of the row vector r^T and X^TX. Therefore, by definition, $\lambda^T G X^T X = \lambda^T \implies \lambda^T \in \mathcal{R}(X^TX)$. We have proved both directions of our claim. Thus, $\lambda^T \in \mathcal{R}(X^TX) \iff \lambda^T G X^T X = \lambda^T$. $\square$

Remark. To check whether $\lambda^{\mathrm{T}} \in \mathcal{R}(\boldsymbol{X})$ is often a tedious job in practice but from (a) and (b), we get an equivalent condition for estimability of a linear function, i.e., by checking whether $\lambda^{\mathrm{T}} G \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} = \lambda^{\mathrm{T}}$.

2. We have the Gram-Schmidt algorithm for a full rank matrix, $X = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$ leading the to the projection matrix ZZ^T as

Gram- Schmidt Algorithm

$X = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$

$\mathbf{z}_1 = \mathbf{x}_1 / \|\mathbf{x}_1\|$

For j from 2 to k

{

$\mathbf{w}_j = \mathbf{x}_j - \left[(\mathbf{z}_1^T \mathbf{x}_j)\mathbf{z}_1 + \ldots (\mathbf{z}_{(j-1)}^T \mathbf{x}_j)\mathbf{z}_{(j-1)} \right]$

$\mathbf{z}_j = \mathbf{w}_j / \|\mathbf{w}_j\|$

}

$Z = [\mathbf{z}_1, \ldots, \mathbf{z}_k]$

Explain how to modify this algorithm when X is not of full rank. Is this case how many columns does Z have?

Assume X is not full rank, and the vector $\mathbf{x}_i$ in the original set of vectors is a linearly dependent sequence. In this case, the corresponding vector $\mathbf{z}_i = 0$. To produce an orthonormal basis for X, we must modify the algorithm to check for 0-vectors in Z and remove them. Specifically, we check whether $\|\mathbf{w}_j\| = 0$. If this condition is true, normalization is impossible, so we skip this vector and proceed to $\mathbf{w}_{j+1}$. The number of columns in Z is equal to the number of columns in X minus the number of linearly dependent vectors in X. In other words, it equals the rank of X.

3. Consider the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

with $\mathbb{E}(\boldsymbol{\varepsilon}) = 0$ and $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I$. X may not have full rank. Here you can use the fact that if $\lambda^T\beta$ is estimable then the OLS estimate is $\widehat{\lambda_1^{\mathrm{T}}\beta} = \mathbf{p}^T M\mathbf{y}$ where $\mathbf{p}$ is the vector associated with estimability and M is the projection matrix onto $\mathcal{W}_X$. Note that based on the introductory remark above you can use a G-inverse in an expression for M.

Suppose $\lambda_1^{\mathrm{T}}\boldsymbol{\beta}$ and $\lambda_2^{\mathrm{T}}\boldsymbol{\beta}$ are both estimable functions. Show that $\text{Cov}[\widehat{\lambda_1^{\mathrm{T}}\boldsymbol{\beta}}, \widehat{\lambda_2^{\mathrm{T}}\boldsymbol{\beta}}] = \sigma^2\lambda_1^{\mathrm{T}}\boldsymbol{G}\lambda_2$, where $\boldsymbol{G} = (X^T X)^-$ is any generalized inverse.

Proof. $\widehat{\lambda_1^T\beta}$ and $\widehat{\lambda_2^T\beta}$ are estimable functions, so they can be written as $c_1^T X\widehat{\beta}$ and $c_2^T X\widehat{\beta}$, respectively. We can substitute these into the covariance expression.

$$\text{Cov}(\widehat{\lambda_1^T\beta}, \widehat{\lambda_2^T\beta}) = \text{Cov}(c_1^T X\widehat{\beta}), c_2^T X\widehat{\beta})$$

We can extract the leading terms, c_1^T and c_2^T.

$$\text{Cov}(c_1^T X\widehat{\beta}, c_2^T X\widehat{\beta}) = c_1^T\text{Cov}(X\widehat{\beta}, X\widehat{\beta})c_2 = c_1^T\text{Var}(X\widehat{\beta})c_2$$

We know $\widehat{\beta} = (X^T X)^- X^T y$. Additionally, we know that $M = X(X^T X)X^T$. Therefore, we can rewrite the expression above as

$$\begin{aligned} c_1^T\text{Var}(X\widehat{\beta})c_2 &= c_1^T\text{Var}(My)c_2 \\ &= c_1^T M\text{Var}(y)M^T c_2 \\ &= \sigma^2 c_1^T M c_2 \end{aligned}$$

We expand M to find

$$\begin{aligned} \sigma^2 c_1^T M c_2 &= \sigma^2(c_1^T X)(X^T X)^-(X^T c_2) \\ &= \sigma^2\lambda_1^T(X^T X)^-(\lambda_2^T)^T \\ &= \sigma^2\lambda_1^T G\lambda_2 \end{aligned}$$

Therefore, $\text{Cov}(\widehat{\lambda_1^T\beta}, \widehat{\lambda_2^T\beta}) = \sigma^2\lambda_1^T G\lambda_2$. $\square$

4. If Rank($\boldsymbol{X}$) has full rank, p, then we know $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$ starting from the normal equations as described in the class and we also know $\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ is an unbiased estimator. Show:

   (a) $\text{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$.

   Proof. We can expand $\text{Cov}(\hat{\beta})$ using the definition of $\hat{\beta}$ as defined above. Additionally, let $A = (X^T X)^{-1}X^T$. We have

$$
\begin{aligned}
\mathrm{Cov}(\hat{\beta}) &= \mathrm{Cov}(Ay, Ay) \\
&= A\mathrm{Cov}(y, y)A^T
\end{aligned}
$$

We know that $\mathrm{Cov}(y, y) = \mathrm{Var}(y, y) = \sigma^2 I$, so the expression simply becomes

$$
A\mathrm{Cov}(y, y)A^T = \sigma^2 AA^T
$$

We can expand again using our definition of A.

$$
\begin{aligned}
\sigma^2 AA^T &= \sigma^2 (X^T X)^{-1} X^T [(X^T X)^{-1} X^T]^T \\
&= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}
$$

Hence, $\mathrm{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. □

(b) $\lambda^{\mathrm{T}} \boldsymbol{\beta}$ is estimable for any $\lambda^{\mathrm{T}} \in \mathbb{R}^p$.

Proof. By Corollary 2.1.10 in Christensen, we know that $\lambda^T \beta$ is estimable if and only if there exists some $c \in \mathbb{R}^p$ such that $\lambda^T \beta = \mathbb{E}[c^T Y]$ for any β. We can expand the RHS of the equation to find

$$
\begin{aligned}
\lambda^T \beta &= \mathbb{E}[c^T Y] \\
&= \mathbb{E}[c^T Y] \\
&= c^T \mathbb{E}[Y] \\
= c^T X \beta \implies \lambda^T &= c^T X
\end{aligned}
$$

It follows that if $\lambda^T = c^T X$ for some $c^T \in \mathbb{R}^p$, then $\lambda^T \beta$ is estimable. Note that since X has full rank, then $\mathcal{R}(X) = \mathbb{R}^p$. Hence, any $\lambda^T \in \mathbb{R}^p$ can be uniquely expressed as a linear combination of the rows of X. In particular, there exists a unique c^T such that $\lambda^T \beta = c^T X$ which implies that $\lambda^T \in \mathcal{R}(X)$. $\lambda^T \beta$ is estimable for all $\lambda^T \in \mathbb{R}^p$. □

(c) $\mathrm{Var}(\lambda^{\mathrm{T}} \widehat{\boldsymbol{\beta}}) = \sigma^2 \lambda^{\mathrm{T}} (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X})^{-1} \lambda$, for any $\lambda^{\mathrm{T}} \in \mathbb{R}^p$.

Proof. We can expand the term $\mathrm{Var}(\lambda^T \widehat{\beta})$ using our definition of $\widehat{\beta}$.

$$
\mathrm{Var}(\lambda^T \widehat{\beta}) = \mathrm{Var}(\lambda^T (X^T X)^{-1} X^T y)
$$

Let $A = \lambda^T (X^T X)^{-1} X^T$. We have

$$
\begin{aligned}
\mathrm{Var}(\lambda^T(X^TX)^{-1}X^Ty) &= \mathrm{Var}(Ay) \\
&= A\mathrm{Var}(y)A^T \\
&= \sigma^2 AA^T \\
&= \sigma^2\lambda^T(X^TX)^{-1}X^T(\lambda^T(X^TX)^{-1}X^T)^T \\
&= \sigma^2\lambda^T(X^TX)^{-1}(X^TX)(X^TX)^{-1}\lambda \\
&= \sigma^2\lambda^T(X^TX)^{-1}\lambda
\end{aligned}
$$

Hence, $\mathrm{Var}(\lambda^T\widehat{\beta}) = \sigma^2\lambda^T(X^TX)^{-1}\lambda$ for any $\lambda^T \in \mathbb{R}^p$. $\square$

(d) $\mathrm{Cov}(\boldsymbol{c_1}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}, \boldsymbol{c_2}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{c_1}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{c_2}$, for any $\boldsymbol{c_1}^{\mathrm{T}}, \boldsymbol{c_2}^{\mathrm{T}} \in \mathbb{R}^p$.

Proof. We can extract the leading terms, namely c_1^T and c_2^T, out of the covariance expression.

$$
\mathrm{Cov}(c_1^T\widehat{\beta}, c_2^T\widehat{\beta}) = c_1^T\mathrm{Cov}(\widehat{\beta}, \widehat{\beta})c_2 = c_1^T\mathrm{Var}(\widehat{\beta})c_2
$$

We again expand $\widehat{\beta} = Ay$, where $A = (X^TX)^{-1}X^T$ for easier simplification.

$$
\begin{aligned}
c_1^T\mathrm{Var}(\widehat{\beta})c_2 &= c_1^T\mathrm{Var}(Ay)c_2 \\
&= c_1^TA\mathrm{Var}(y)A^Tc_2 \\
&= \sigma^2 c_1^TAA^Tc_2 \\
&= \sigma^2 c_1^T(X^TX)^{-1}(X^TX)(X^TX)^{-1}c_2 \\
&= \sigma^2 c_1^T(X^TX)^{-1}c_2
\end{aligned}
$$

Hence, for any $c_1^T, c_2^T \in \mathbb{R}^p$, we have that $\mathrm{Cov}(c_1^T\widehat{\beta}, c_2^T\widehat{\beta}) = \sigma^2 c_1^T(X^TX)^{-1}c_2$