The Orthogonal Projection Matrix onto $\mathcal{C}(\mathbf{X})$ and the Normal Equations
Theorem 2.1:

$P_X$, the orthogonal projection matrix onto $\mathcal{C}(X)$, is equal to $X(X'X)^{-1}X'$. The matrix $X(X'X)^{-1}X'$ satisfies

(a) $[X(X'X)^{-1}X'][X(X'X)^{-1}X'] = [X(X'X)^{-1}X']$

(b) $X(X'X)^{-1}X'y \in \mathcal{C}(X) \ \forall \ y \in \mathbb{R}^n$

(c) $X(X'X)^{-1}X'x = x \ \forall \ x \in \mathcal{C}(X)$

(d) $X(X'X)^{-1}X' = [X(X'X)^{-1}X']'$

(e) $X(X'X)^{-1}X'$ is the same $\forall$ Gl of $X'X$. 

Copyright ©2012 Dan Nettleton (Iowa State University)
It is useful to establish a few results before proving Theorem 2.1.

Show that if $A$ is a symmetric matrix and $G$ is a GI of $A$, then $G'$ is also GI of $A$. 
Proof:

\[ AGA = A \Rightarrow (AGA)' = A' \]

\[ \Rightarrow A'G'A' = A' \]

\[ \Rightarrow AG'A = A \quad (\because A = A') \]

\[ \therefore G' \text{ is a GI of } A. \]
Result 2.4:

\[ X'XA = X'XB \iff XA = XB. \]

One direction is obvious (\(\iff\)).

Can you prove the other (\(\implies\))?
Proof (\(\iff\)):

\[
(XA - XB)'(XA - XB) = (A'X' - B'X')(XA - XB)
\]

\[
= A'X'XA - A'X'XB - B'X'XA + B'X'XB
\]

\[
= A'X'XA - A'X'XA - B'X'XA + B'X'XA
\]

\[
= 0.
\]
\[ X_A - X_B = 0 \Rightarrow X_A = X_B. \]
Result 2.5:

Suppose \((X'X)^{-}\) is any GI of \(X'X\). Then \((X'X)^{-}X'\) is a GI of \(X\), i.e.,

\[
X(X'X)^{-}X'X = X.
\]
Proof of Result 2.5

Since \((X'X)^{-}\) is a GI of \(X'X\),

\[ X'X(X'X)^{-}X'X = X'X. \]

By Result 2.4, the result follows.

(Take \(A = (X'X)^{-}X'X\) and \(B = I\). Then
\[XA = X(X'X)^{-}X'X = X = XB.\)\]

\[\Box\]
Corollary to Result 2.5:

For \((X'X)^{-}\) any GI of \(X'X\),

\[
X'X(X'X)^{-}X' = X'.
\]
Proof of Corollary:

Suppose \((X'X)^-\) is any GI of \(X'X\).

\[ \therefore X'X \text{ is symmetric, } [(X'X)^-]' \text{ is also a GI of } X'X. \]
Thus Result 2.5 implies that

\[ X[(X'X)^{-}]'X'X = X \]

\[ \Rightarrow [X[(X'X)^{-}]'X'X]' = [X]' \]

\[ \Rightarrow X'X(X'X)^{-}X' = X'. \]
Now we can prove Theorem 2.1:

First show $X(X'X)^{-}X'$ is idempotent.
\[ X(X'X)^{-}X'X(X'X)^{-}X' = X(X'X)^{-}X' \] by Result 2.5.
Now show that

\[ X(X'X)^{-1}X'y \in C(X) \forall y \in \mathbb{R}^n. \]
\[ X(X'X)^{-1}X'y = Xz, \text{ where } z = (X'X)^{-1}X'y. \]

Thus, \( X(X'X)^{-1}X'y \in C(X). \)
Now show that

\[ X(X'X)^{-}X'z = z \quad \forall \ z \in C(X). \]
If $z \in C(X)$, $\exists c \ni z = Xc$. ∴

$$X(X'X)^{-1}X'z = X(X'X)^{-1}X'Xc$$

$$= Xc \text{ (By Result 2.5)}$$

$$= z.$$  

Thus, we have $X(X'X)^{-1}X'z = z \forall z \in C(X)$. 
Now we know $X(X'X)^{-1}X'$ is a projection matrix onto $C(X)$.

To be the unique projection matrix onto $C(X)$, $X(X'X)^{-1}X'$ must be symmetric.
To show symmetry of $X(X'X)^{-}X'$, it will help to first show that

$$X(X'X)_1^{-}X' = X(X'X)_2^{-}X'$$

for any two GIs of $X'X$ denoted $(X'X)_1^{-}$ and $(X'X)_2^{-}$. 
By the Corollary to Result 2.5

\[ X' = X'X(X'X)^{-1}X'. \]

Therefore,

\[ X(X'X)^{-1}X' = X(X'X)^{-1}X'X(X'X)^{-1}X' \]
\[ = X(X'X)^{-1}X' \] by Result 2.5.

\[ \therefore X(X'X)^{-1}X' \] is the same regardless of which GL for \( X'X \) is used.
Now show that $X(X'X)^{-}X'$ is symmetric.
\[
X(X'X)^{-}X' = X[(X'X)^{-}]'X' \\
= X(X'X)^{-}X'
\]

\[\because\] symmetry of \(X'X\) implies that \([(X'X)^{-}]'\) is a GI of \(X'X\), and \(X(X'X)^{-}X'\) is the same regardless of choice of GI.
We have shown that $X(X'X)^{-1}X'$ is the orthogonal projection matrix onto $C(X)$.

We know that it is the only symmetric projection matrix onto $C(X)$ by Result A.16.
Example:

Find the orthogonal projection matrix onto $\mathcal{C}(\begin{pmatrix} 1 \end{pmatrix})$.

Also, simplify $P_{Xy}$ in this case.
\(X'X = 1'1 = n. \) Thus \((X'X)^{-} = \frac{1}{n}.\)

\[
X(X'X)^{-}X' = 1 \begin{bmatrix} \frac{1}{n} \end{bmatrix} 1' \\
= \frac{1}{n} 11'.
\]
Thus, when $X = 1$, $P_X$ is an $n \times n$ matrix whose entries are each $\frac{1}{n}$.

We saw previously the special case $\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$.
\[ P_{xy} = \frac{1}{n} 11' y \]

\[ = \frac{1}{n} 1 \sum_{i=1}^{n} y_i = 1 \frac{1}{n} \sum_{i=1}^{n} y_i \]

\[ = 1 \bar{y} = \begin{bmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix}. \]
By Corollary A.4 to Result A.16, we know that

$$I - P_X = I - X(X'X)^{-1}X'$$

is the orthogonal projection matrix onto $C(X)\perp = \mathcal{N}(X')$.

This is Result 2.6.
Fitted Values and Residuals

By Result A.4, any \( y \in \mathbb{R}^n \) may be written as

\[
y = s + t,
\]

where \( s \in C(X) \) and \( t \in C(X)^\perp = \mathcal{N}(X') \).

Furthermore, the vectors \( s \in C(X) \) and \( t \in C(X)^\perp \) are unique.
Fitted Values and Residuals

Because

\[ y = P_X y + (I - P_X) y \]

\[ P_X y \in C(X) \text{ and } (I - P_X) y \in C(X)^\perp, \]

we get \( s = P_X y \) and \( t = (I - P_X) y \).

Moreover,

\[ \hat{y} \equiv P_X y = \text{ the vector of fitted values} \]

\[ \hat{e} \equiv (I - P_X) y = y - \hat{y} = \text{ the vector of residuals.} \]
Let $P_W$ and $P_X$ denote the orthogonal projection matrices onto $C(W)$ and $C(X)$, respectively.

Suppose $C(W) \subseteq C(X)$. Show that

$$P_W P_X = P_X P_W = P_W.$$
Proof:

\[ C(W) \subseteq C(X) \Rightarrow \exists \ B \ni XB = W. \]

\[
\begin{align*}
\therefore \ P_XP_W &= X(X'X)^{-1}X'W(W'W)^{-1}W' \\
&= X(X'X)^{-1}X'XB(W'W)^{-1}W' \\
&= XB(W'W)^{-1}W' \\
&= W(W'W)^{-1}W' \\
&= P_W.
\end{align*}
\]
Now

\[ P_X P_W = P_W \Rightarrow (P_X P_W)' = P'_W \]

\[ \Rightarrow P'_W P'_X = P'_W \]

\[ \Rightarrow P_W P_X = P_W. \]
Theorem 2.2:

If $C(W) \subseteq C(X)$, then $P_X - P_W$ is the orthogonal projection matrix onto $C((I - P_W)X)$. 
Proof of Theorem 2.2:

\[(P_X - P_W)' = P'_X - P'_W = P_X - P_W.\]

\[\therefore P_X - P_W \text{ is symmetric.}\]

\[(P_X - P_W)(P_X - P_W) = P_XP_X - P_WP_X - P_XP_W + P_WP_W \]
\[= P_X - P_W - P_W + P_W \]
\[= P_X - P_W.\]

\[\therefore P_X - P_W \text{ is idempotent.}\]
Is \((P_X - P_W)y \in C((I - P_W)X) \forall y?\)

\[
(P_X - P_W)y = (P_X - P_W P_X)y
= (I - P_W) P_X y
= (I - P_W) X (X' X)^{-1} X' y
\in C((I - P_W)X).
\]
Is \((P_X - P_W)z = z\) \(\quad z \in \mathcal{C}((I - P_W)X)\)?

\[
(P_X - P_W)(I - P_W)X = P_X(I - P_W)X - P_W(I - P_W)X
\]
\[
= (P_X - P_XP_W)X - (P_W - P_WP_W)X
\]
\[
= (P_X - P_W)X - (P_W - P_W)X
\]
\[
= (P_X - P_W)X
\]
\[
= (I - P_W)P_XX
\]
\[
= (I - P_W)X.
\]
Now $z \in C((I - P_W)X)$.

$\Rightarrow z = (I - P_W)Xc$ for some $c$. Therefore,

$$(P_X - P_W)z = (P_X - P_W)(I - P_W)Xc$$

$$= (I - P_W)Xc$$

$$= z.$$
We have previously seen that

\[ Q(b) = (y - Xb)'(y - Xb) = \|y - Xb\|^2 \geq \|y - P_Xy\|^2 \]

\( \forall \ b \in \mathbb{R}^p \) with equality iff \( Xb = P_Xy \).
Now we know that $P_X = X(X'X)^{-1} X'$.

Thus, $\hat{\beta}$ minimizes $Q(b)$ iff $X\hat{\beta} = X(X'X)^{-1} X'y$.

By Result 2.4, this equation is equivalent to

$$X'X\hat{\beta} = X'X(X'X)^{-1} X'y.$$
Because $X'X(X'X)^{-1}X' = X'$, $X'X\hat{\beta} = X'X(X'X)^{-1}X'y$ is equivalent to

$$X'X\hat{\beta} = X'y.$$ 

This system of linear equations is known as the **Normal Equations (NE)**.
We have established Result 2.3:

\( \hat{\beta} \) is a solution to the NE \( (X'Xb = X'y) \) iff \( \hat{\beta} \) minimizes \( Q(b) \).
Corollary 2.1:

The NE are consistent.
Proof of Corollary 2.1:

NE are $X'Xb = X'y$.

If we take $\hat{\beta} = (X'X)^{-1}X'y$, then

$$X'X\hat{\beta} = X'X(X'X)^{-1}X'y$$

$$= X'y.$$
By Result A.13, \( \hat{\beta} \) is a solution to \( X'Xb = X'y \) iff

\[
\hat{\beta} = (X'X)^{-1}X'y + [I - (X'X)^{-1}X'X]z
\]

for some \( z \in \mathbb{R}^p \).
Corollary 2.3:

$X\hat{\beta}$ is invariant to the choice of a solution $\hat{\beta}$ to the NE, i.e., if $\hat{\beta}_1$ and $\hat{\beta}_2$ are any two solutions to the NE, then $X\hat{\beta}_1 = X\hat{\beta}_2$. 
Proof of Corollary 2.3:

\[ X'X\hat{\beta}_1 = X'X\hat{\beta}_2 (= X'y) \]
\[ \Rightarrow X\hat{\beta}_1 = X\hat{\beta}_2 \text{ by Result 2.4.} \]
We will finish this set of notes with some other results from Section 2.2 of the text.
Lemma 2.1: $\mathcal{N}(X'X) = \mathcal{N}(X)$.

Result 2.2: $\mathcal{C}(X'X) = \mathcal{C}(X')$.

Corollary 2.2: $\text{rank}(X'X) = \text{rank}(X)$.
Proof of Lemma 2.1:

\[ Xc = 0 \Rightarrow X'Xc = 0 \]

\[ \therefore \mathcal{N}(X) \subseteq \mathcal{N}(X'X). \]

\[ X'Xc = 0 \Rightarrow c'X'Xc = 0 \]

\[ \Rightarrow Xc = 0 \]

\[ \therefore \mathcal{N}(X'X) \subseteq \mathcal{N}(X). \]

Thus, \( \mathcal{N}(X'X) = \mathcal{N}(X) \).
Proof of Result 2.2:

\[ X'Xc = X'(Xc) \Rightarrow C(X'X) \subseteq C(X'). \]

\[ X'c = X'P_xc = X'X(X'X)^{-1}X'c \]
\[ = X'X[(X'X)^{-1}X'c] \]
\[ \Rightarrow C(X') \subseteq C(X'X). \]

Thus, \( C(X'X) = C(X'). \) \( \square \)
Proof of Corollary 2.2:

\[ C(X'X) = C(X') \]
\[ \Rightarrow \dim(C(X'X)) = \dim(C(X')) \]
\[ \Rightarrow \text{rank}(X'X) = \text{rank}(X') \]
\[ \Rightarrow \text{rank}(X'X) = \text{rank}(X). \]

(Can facilitate finding \((X'X)^{-1}\) needed for \(\hat{\beta}\) and \(P_X\).)