1. (a) Let $(\boldsymbol{X}'\boldsymbol{X})^-$ be any generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$, which by definition implies

$$\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{X},$$

where $\boldsymbol{X}$ has dimension $m \times n$, say. Put $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{B} = \boldsymbol{I}_{n\times n}$, so that

$$\boldsymbol{X}'\boldsymbol{X}\underbrace{(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}}_{\boldsymbol{A}} = \boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{X}\underbrace{\boldsymbol{I}_{n\times n}}_{\boldsymbol{B}} \implies \boldsymbol{X}'\boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{B}.$$

By the result of problem 7 on Homework 1, it follows that $\boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}\boldsymbol{B}$, and hence

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}.$$

(b) Let $\boldsymbol{A}$ be any symmetric matrix and $\boldsymbol{G}$ be any generalized inverse of $\boldsymbol{A}$. By definition,

$$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}.$$

Now, transpose both sides and use the fact that $\boldsymbol{A}' = \boldsymbol{A}$ by symmetry:

$$\begin{aligned}(\boldsymbol{A}\boldsymbol{G}\boldsymbol{A})' = \boldsymbol{A}' &\implies \boldsymbol{A}'\boldsymbol{G}'\boldsymbol{A}' = \boldsymbol{A}' \\ &\implies \boldsymbol{A}\boldsymbol{G}'\boldsymbol{A} = \boldsymbol{A}.\end{aligned}$$

Hence, $\boldsymbol{G}'$ is a generalized inverse of $\boldsymbol{A}$.

(c) Let $\boldsymbol{G}$ be any generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$. Notice that $\boldsymbol{X}'\boldsymbol{X}$ is symmetric, so by part (b), $\boldsymbol{G}'$ is also a generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$. The result of part (a) holds for any generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$, and hence holds using $\boldsymbol{G}'$. Using the result of part (a) with $\boldsymbol{G}'$ and then taking transposes gives

$$\begin{aligned}\boldsymbol{X}\boldsymbol{G}'\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X} &\implies (\boldsymbol{X}\boldsymbol{G}'\boldsymbol{X}'\boldsymbol{X})' = \boldsymbol{X}' \\ &\implies \boldsymbol{X}'[\boldsymbol{X}']'[\boldsymbol{G}']'\boldsymbol{X}' = \boldsymbol{X}' \\ &\implies \boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}' = \boldsymbol{X}'.\end{aligned}$$

Because we chose $\boldsymbol{G}$ to be any generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$,

$$\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \boldsymbol{X}'.$$

Comments: We could have

$$(\boldsymbol{X}'\boldsymbol{X})^- \neq [(\boldsymbol{X}'\boldsymbol{X})^-]',$$

so it is important that (a) and (b) are used at the right steps in your proof so it is clear that you aren't trying to say $(\boldsymbol{X}'\boldsymbol{X})^- = [(\boldsymbol{X}'\boldsymbol{X})^-]'$. On a related note, we may also have $[(\boldsymbol{X}'\boldsymbol{X})^-]' \neq [(\boldsymbol{X}'\boldsymbol{X})']^-$.

(d) This part requires *two* proofs that $\boldsymbol{P_X}$ is idempotent for full credit.

1. By part (a),

$$\begin{aligned}\boldsymbol{P_X}\boldsymbol{P_X} &= \underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}}_{\boldsymbol{X}}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' \\ &= \boldsymbol{P_X}.\end{aligned}$$

2. By part (c),

$$\begin{aligned}\boldsymbol{P_X}\boldsymbol{P_X} &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\underbrace{\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'}_{\boldsymbol{X}'} \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' \\ &= \boldsymbol{P_X}.\end{aligned}$$

(e) Let $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ be any generalized inverses of $\boldsymbol{X}'\boldsymbol{X}$. By parts (a) and (c), we have

$$\begin{aligned}\boldsymbol{X}\boldsymbol{G}_1\boldsymbol{X}' &= \boldsymbol{X}\boldsymbol{G}_1\underbrace{\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}_2\boldsymbol{X}'}_{\boldsymbol{X}'} && \text{part (c) holds for any generalized inverse of } \boldsymbol{X}'\boldsymbol{X} \\ &= \underbrace{\boldsymbol{X}\boldsymbol{G}_1\boldsymbol{X}'\boldsymbol{X}}_{\boldsymbol{X}}\boldsymbol{G}_2\boldsymbol{X}' && \text{part (a) holds for any generalized inverse of } \boldsymbol{X}'\boldsymbol{X} \\ &= \boldsymbol{X}\boldsymbol{G}_2\boldsymbol{X}'.\end{aligned}$$

Comments:

- A few students tried to use the fact that $\boldsymbol{P_X}$ is the same matrix regardless of which generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$ is used, but this is what we are trying to show.
- This statement should hold for any two generalized inverse matrices of $\boldsymbol{X}'\boldsymbol{X}$. Some students proved this by setting $\boldsymbol{G}_2 = (\boldsymbol{G}_1)'$. This case cannot generalize this result.

(f) Let $(\boldsymbol{X}'\boldsymbol{X})^-$ be any generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$. We know that $\boldsymbol{X}'\boldsymbol{X}$ is a symmetric matrix, so the result of part (b) says that if $(\boldsymbol{X}'\boldsymbol{X})^-$ is a generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$, then $[(\boldsymbol{X}'\boldsymbol{X})^-]'$ is a generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$. The result of part (e) then establishes that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \boldsymbol{X}[(\boldsymbol{X}'\boldsymbol{X})^-]'\boldsymbol{X}'$. Hence, these results and properties of matrix transpose give

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{X}}' &= \left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\right)' \\ &= [\boldsymbol{X}']'[(\boldsymbol{X}'\boldsymbol{X})^-]'\boldsymbol{X}' \\ &= \boldsymbol{X}[(\boldsymbol{X}'\boldsymbol{X})^-]'\boldsymbol{X}' \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' && \text{by parts (d,g) as explained above} \\ &= \boldsymbol{P_X}.\end{aligned}$$

Comments: It is important to use parts (d) and (g) at the right steps in your proof so it is clear that you aren't trying to say $(\boldsymbol{X}'\boldsymbol{X})^- = [(\boldsymbol{X}'\boldsymbol{X})^-]'$.

2. Let $\boldsymbol{X}$ be an $n \times p$ matrix and $\boldsymbol{y}$ be an $n \times 1$ vector. Suppose that $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X})$ and $\boldsymbol{z} \neq \boldsymbol{P_X y}$, which implies $(\boldsymbol{P_X y} - \boldsymbol{z}) \neq \boldsymbol{0}_{n\times 1}$. Observe that $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X})$ implies that $\boldsymbol{P_X z} = \boldsymbol{z}$. Using this result and the fact that $\boldsymbol{P_X}$ is symmetric and idempotent, it follows that

$$
\begin{aligned}
(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) &= (\boldsymbol{y}' - [\boldsymbol{P_X y}]')(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= (\boldsymbol{y}' - \boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{X}}')(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= (\boldsymbol{y}' - \boldsymbol{y}'\boldsymbol{P_X})(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= \boldsymbol{y}'\boldsymbol{P_X y} - \boldsymbol{y}'\boldsymbol{z} - \boldsymbol{y}'\boldsymbol{P_X P_X y} + \boldsymbol{y}'\boldsymbol{P_X z} \\
&= \boldsymbol{y}'\boldsymbol{P_X y} - \boldsymbol{y}'\boldsymbol{z} - \boldsymbol{y}'\boldsymbol{P_X y} + \boldsymbol{y}'\boldsymbol{P_X z} \\
&= -\boldsymbol{y}'\boldsymbol{z} + \boldsymbol{y}'\boldsymbol{z} \\
&= 0.
\end{aligned}
$$

Now that we have $(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) = 0$ and $(\boldsymbol{P_X y} - \boldsymbol{z}) \neq \boldsymbol{0}$, we can use the same argument provided in the homework with $\boldsymbol{a} = \boldsymbol{y} - \boldsymbol{P_X y}$ and $\boldsymbol{b} = \boldsymbol{P_X y} - \boldsymbol{z}$:

$$
\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{z}\|^2 &= \|\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z}\|^2 \\
&= (\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z})'\,(\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z}) \\
&= \big((\boldsymbol{y} - \boldsymbol{P_X y})' + (\boldsymbol{P_X y} - \boldsymbol{z})'\big)\big((\boldsymbol{y} - \boldsymbol{P_X y}) + (\boldsymbol{P_X y} - \boldsymbol{z})\big) \\
&= (\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{y} - \boldsymbol{P_X y}) + 2(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) + (\boldsymbol{P_X y} - \boldsymbol{z})'(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= \|\boldsymbol{y} - \boldsymbol{P_X y}\|^2 + \|\boldsymbol{P_X y} - \boldsymbol{z}\|^2 \\
&> \|\boldsymbol{y} - \boldsymbol{P_X y}\|^2 .
\end{aligned}
$$

Hence, $\|\boldsymbol{y} - \boldsymbol{z}\| > \|\boldsymbol{y} - \boldsymbol{P_X y}\|$, which says that $\boldsymbol{P_X y}$ is the unique point in $\mathcal{C}(\boldsymbol{X})$ that is closest to $\boldsymbol{y}$ in Euclidean distance.

Comments: You can instead show that $(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) = 0$ by orthogonality, but as this is a proof, you need to provide sufficient reasoning or work to establish this.

3. Key:
1. $\underset{n\times 1}{\boldsymbol{a}} \in \mathcal{C}(\boldsymbol{X}) \iff \boldsymbol{a} = \boldsymbol{X} \underset{p\times 1}{\boldsymbol{b}}$ for some $\boldsymbol{b}$
2. $\boldsymbol{P_X X} = \boldsymbol{X}$ by property of projection matrix

Prove that $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{P_X})$:

$$
\begin{aligned}
\boldsymbol{a} \in \mathcal{C}(\boldsymbol{X}) &\iff \boldsymbol{a} = \boldsymbol{Xb} && \text{for some } \boldsymbol{b} \text{ by key 1} \\
&\iff \boldsymbol{a} = \underbrace{\boldsymbol{P_X X}}_{\boldsymbol{X}}\boldsymbol{b} && \text{for some } \boldsymbol{b} \text{ by key 2} \\
&\iff \boldsymbol{a} = \boldsymbol{P_X}\underbrace{\boldsymbol{Xb}}_{n\times 1} && \text{treat as } \boldsymbol{P_X} \text{ product a } n \times 1 \text{ vector} \\
&\iff \boldsymbol{a} = \boldsymbol{P_X k} && \text{for some } \boldsymbol{k} = \boldsymbol{Xb} \\
&\implies \boldsymbol{a} \in \mathcal{C}(\boldsymbol{P_X}) && \text{by key 1}
\end{aligned}
$$

So $\mathcal{C}(\boldsymbol{X}) \subseteq \mathcal{C}(\boldsymbol{P_X})$.

Then similarly,

$$
\begin{aligned}
\boldsymbol{g} \in \mathcal{C}(\boldsymbol{P_X}) &\iff \boldsymbol{g} = \boldsymbol{P_X h} && \text{for some } n \times 1 \text{ vector } \boldsymbol{h} \text{ by key 1} \\
&\iff \boldsymbol{g} = \underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'}_{\boldsymbol{P_X}}\boldsymbol{h} && \text{for some } \boldsymbol{h} \\
&\iff \boldsymbol{g} = \boldsymbol{X}\underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{h}}_{p\times 1} && \text{treat as } \boldsymbol{X} \text{ product a } p \times 1 \text{ vector} \\
&\iff \boldsymbol{g} = \boldsymbol{Xq} && \text{for some } \boldsymbol{q} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{h} \\
&\implies \boldsymbol{g} \in \mathcal{C}(\boldsymbol{X}) && \text{by key 1}
\end{aligned}
$$

So $\mathcal{C}(\boldsymbol{P_X}) \subseteq \mathcal{C}(\boldsymbol{X})$. According to the results above, $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{P_X})$.

4. Prove $(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ is a solution to the normal equations $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$ (by slide 8 of set 2).

   Let $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$:

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} &= \boldsymbol{X}'\underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'}_{\boldsymbol{P_X}}\boldsymbol{y} \\
&= \boldsymbol{X}'\boldsymbol{P_X}\boldsymbol{y} \\
&= \boldsymbol{X}'\boldsymbol{y} && \boldsymbol{X}'\boldsymbol{P_X} = \boldsymbol{X}' \text{ by property of projection matrix in slide 5 of set 2}
\end{aligned}
$$

   Therefore $(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ is a solution to the normal equations.

5. Suppose the Gauss-Markov model with normal errors holds (see slide 16 of slide set 2 for a precise statement of the model).

   (a) Suppose $\boldsymbol{C\beta}$ is estimable. Derive the distribution of $\boldsymbol{C}\hat{\boldsymbol{\beta}}$, the OLSE of $\boldsymbol{C\beta}$.

   $\boldsymbol{C\beta}$ is estimable $\implies$ there exists $\boldsymbol{A}$ that $\boldsymbol{C} = \boldsymbol{AX}$

$$
\begin{aligned}
\boldsymbol{C}\hat{\boldsymbol{\beta}} &= \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} \\
&= \boldsymbol{AX}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} && \boldsymbol{C} = \boldsymbol{AX} \\
&= \boldsymbol{Aprojy} && \boldsymbol{proj} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'
\end{aligned}
$$

   Based on the model assumptions, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$. Then $\boldsymbol{C}\hat{\boldsymbol{\beta}} = \boldsymbol{AP_Xy}$ is also multivariate normal by slide 32 of set 1, $\boldsymbol{AP_Xy} \sim \mathcal{N}(\boldsymbol{AP_XX\beta}, \boldsymbol{AP_X}\sigma^2\boldsymbol{I}(\boldsymbol{AP_X})')$

$$\boldsymbol{AP_XX\beta} = \boldsymbol{AX\beta} = \boldsymbol{C\beta}$$

$$
\begin{aligned}
\boldsymbol{AP_X}\sigma^2\boldsymbol{I}(\boldsymbol{AP_X})' &= \sigma^2\boldsymbol{AP_XP_X'A'} \\
&= \sigma^2\boldsymbol{AP_XA'} && \boldsymbol{P_X} \text{ is symmetric and idempotent} \\
&= \sigma^2\boldsymbol{AX}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{A}' \\
&= \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'
\end{aligned}
$$

   Therefore $\boldsymbol{C}\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{C\beta}, \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')$.

(b) Now suppose $\boldsymbol{C\beta}$ is NOT estimable.

$$\begin{aligned} Var(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}) &= (\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}')\sigma^2\boldsymbol{I}(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}')' \\ &= \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-'}\boldsymbol{C}' \end{aligned}$$

We can not simply this further when $\boldsymbol{C\beta}$ is NOT estimable.

(c) Now suppose $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ is testable. Prove the result on slide 23 of set 2.

Given the hypothesis is testable (see slide 18 of set 2), $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is estimable and from the resluts in part (a), we have $\boldsymbol{c}'\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{c}'\boldsymbol{\beta}, \sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c})$, by linear transformation,

$$\frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}} \sim \mathcal{N}\left(\frac{\boldsymbol{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}, 1\right)$$

let $u = \frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}$ and $\delta = \frac{\boldsymbol{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}$, $u \sim \mathcal{N}(\delta, 1)$.

Then by slide 19 of set 2,

$$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{n-r}}{n-r} \implies w = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$$

$\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent, so u and w, which are functions of $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, respectively, are also independent (see Theorem 4.3.5 in Casella and Berger, 2002).
By slide 39 of set 1,

$$\frac{u}{\sqrt{w/(n-r)}} = \frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\hat{\sigma}^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}} \sim t_{n-r}(\delta)$$

Therefore, it follows a t distribution with non-central parameter $\delta = \frac{\boldsymbol{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}$ and degrees of freedom $n - r$.

Note: The independence between u and w is necessary. We can first show independence of $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$. Because $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is estimable, we can write it as $\boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{a}'\boldsymbol{P_X}\boldsymbol{y}$ for some $\boldsymbol{a}'$, and $\hat{\sigma}^2 = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/(n-r) = ||(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}||^2/(n-r)$.

Now we use **the independence results on slide 38 in set 1**. When $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ in GMMNE (slide 16 of set 2), let $\boldsymbol{A}_1 = \boldsymbol{a}'\boldsymbol{P_X}$, and $\boldsymbol{A}_2 = (\boldsymbol{I} - \boldsymbol{P_X})/(n-r)$. Then

$$\begin{aligned} \boldsymbol{A}_1\sigma^2\boldsymbol{I}\boldsymbol{A}_2' &= \boldsymbol{a}'\boldsymbol{P_X}\sigma^2\boldsymbol{I}(\boldsymbol{I} - \boldsymbol{P_X})'/(n-r) \\ &= \sigma^2\boldsymbol{a}'\boldsymbol{P_X}(\boldsymbol{I} - \boldsymbol{P_X})'/(n-r) \\ &= \sigma^2\boldsymbol{a}'\boldsymbol{P_X}(\boldsymbol{I} - \boldsymbol{P_X})/(n-r) \\ &= \sigma^2\boldsymbol{a}'(\boldsymbol{P_X} - \boldsymbol{P_X}\boldsymbol{P_X})/(n-r) \\ &= \boldsymbol{0} \qquad\qquad \text{because } \boldsymbol{P_X} \text{ is idempotent.} \end{aligned}$$

Then we have $\boldsymbol{c}'\hat{\boldsymbol{\beta}} \perp \hat{\sigma}^2$, which impies $u \perp w$ by Theorem 4.3.5 in Casella and Berger(2002).