ML and REML Variance Component Estimation

Suppose

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

where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ for some positive definite, symmetric matrix $\boldsymbol{\Sigma}$.

Furthermore, suppose each element of $\boldsymbol{\Sigma}$ is a known function of an unknown vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

For example, consider

$$\boldsymbol{\Sigma} = \sum_{j=1}^{m} \sigma_j^2 \boldsymbol{Z}_j \boldsymbol{Z}_j' + \sigma_e^2 \boldsymbol{I},$$

where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are known matrices,

$$\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m, \theta_{m+1}]' = [\sigma_1^2, \ldots, \sigma_m^2, \sigma_e^2]',$$

and $\boldsymbol{\Theta} = \{\boldsymbol{\theta} : \theta_j > 0,\ j = 1, \ldots, m+1\}$.

The likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})}$$

and the log likelihood function is

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\beta)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta).$$

Based on previous results,

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

is minimized over $\boldsymbol{\beta} \in \mathbb{R}^p$, for any fixed $\boldsymbol{\theta}$, by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \equiv (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}.$$

Thus, the profile log likelihood for $\boldsymbol{\theta}$ is

$$\begin{aligned} l^*(\boldsymbol{\theta}|\boldsymbol{y}) &= \sup_{\boldsymbol{\beta}\in\mathbb{R}^p} l(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})). \end{aligned}$$

In the general case, numerical methods are used to find the maximizer of $l^*(\boldsymbol{\theta}|\boldsymbol{y})$ over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Let

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max\{l^*(\boldsymbol{\theta}|\boldsymbol{y}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}.$$

Let $\hat{\boldsymbol{\Sigma}}_{\text{MLE}}$ be the matrix $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ in place of $\boldsymbol{\theta}$.

The MLE of an estimable $\boldsymbol{C\beta}$ is then given by

$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_{\text{MLE}} = \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}_{\text{MLE}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}_{\text{MLE}}^{-1}\boldsymbol{y}.$$

We can write $\boldsymbol{\Sigma}$ as $\sigma^2\boldsymbol{V}$, where $\sigma^2 > 0, \boldsymbol{V}$ is PD and symmetric, σ^2 is known function of $\boldsymbol{\theta}$, and each entry of $\boldsymbol{V}$ is a known function of $\boldsymbol{\theta}$, e.g.,

$$\begin{aligned}\boldsymbol{\Sigma} &= \sum_{j=1}^{m} \sigma_j^2 \boldsymbol{Z}_j \boldsymbol{Z}_j' + \sigma_e^2 \boldsymbol{I} \\ &= \sigma_e^2 \left[\sum_{j=1}^{m} \frac{\sigma_j^2}{\sigma_e^2} \boldsymbol{Z}_j \boldsymbol{Z}_j' + \boldsymbol{I} \right]\end{aligned}$$

If we let $\hat{\sigma}^2_{\text{MLE}}$ and $\hat{\boldsymbol{V}}_{\text{MLE}}$ denote σ^2 and $\boldsymbol{V}$ with $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ in place of $\boldsymbol{\theta}$, then

$$\hat{\boldsymbol{\Sigma}}_{\text{MLE}} = \hat{\sigma}^2_{\text{MLE}} \hat{\boldsymbol{V}}_{\text{MLE}} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}^{-1}_{\text{MLE}} = \frac{1}{\hat{\sigma}^2_{\text{MLE}}} \hat{\boldsymbol{V}}^{-1}_{\text{MLE}}.$$

It follows that

$$\begin{aligned}
\boldsymbol{C}\hat{\boldsymbol{\beta}}_{\text{MLE}} &= \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}_{\text{MLE}}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}_{\text{MLE}}\boldsymbol{y} \\
&= \boldsymbol{C}\left(\boldsymbol{X}'\frac{1}{\hat{\sigma}^2_{\text{MLE}}}\hat{\boldsymbol{V}}^{-1}_{\text{MLE}}\boldsymbol{X}\right)^{-}\boldsymbol{X}'\frac{1}{\hat{\sigma}^2_{\text{MLE}}}\hat{\boldsymbol{V}}^{-1}_{\text{MLE}}\boldsymbol{y} \\
&= \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}_{\text{MLE}}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}_{\text{MLE}}\boldsymbol{y}.
\end{aligned}$$

Note that

$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_{\text{MLE}} = \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{V}}_{\text{MLE}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{V}}_{\text{MLE}}^{-1}\boldsymbol{y}$$

is the GLS estimator under the Aitken model with $\hat{\boldsymbol{V}}_{\text{MLE}}$ in place of $\boldsymbol{V}$.

- We have seen by example that the MLE of the variance component vector $\boldsymbol{\theta}$ can be biased.

- For example, for the case of $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$, where $\boldsymbol{\theta} = [\sigma^2]$, the MLE of σ^2 is

$$\frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n} \text{ with expectation } \frac{n-r}{n}\sigma^2.$$

This MLE for σ^2 is often criticized for "failing to account for the loss of degrees of freedom needed to estimate $\boldsymbol{\beta}$."

$$\begin{aligned} E\left[\frac{(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{n}\right] &= \frac{n-r}{n}\sigma^2 \\ &< \sigma^2 \\ &= E\left[\frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}{n}\right] \end{aligned}$$

A familiar special case: Suppose

$$y_1, \ldots, y_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2).$$

Then

$$E\left[\frac{\sum_{i=1}^n (y_i - \mu)^2}{n}\right] = \sigma^2, \text{ but}$$

$$E\left[\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}\right] = \frac{n-1}{n}\sigma^2.$$

- REML is an approach that produces unbiased estimators for these special cases and produces less biased estimators than ML estimators in general.

- Depending on whom you ask, REML stands for REsidual Maximum Likelihood or REstricted Maximum Likelihood.

The REML method:

- Find $n - rank(\boldsymbol{X}) = n - r$ linearly independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}$ such that $\boldsymbol{a}_i'\boldsymbol{X} = \boldsymbol{0}'$ for all $i = 1, \ldots, n - r$.

- Find the maximum likelihood estimate of $\boldsymbol{\theta}$ using $w_1 \equiv \boldsymbol{a}_1'\boldsymbol{y}, \ldots, w_{n-r} \equiv \boldsymbol{a}_{n-r}'\boldsymbol{y}$ as data.

$$A = [a_1, \ldots, a_{n-r}] \qquad w = \begin{bmatrix} w_1 \\ \vdots \\ w_{n-r} \end{bmatrix} = \begin{bmatrix} a_1'y \\ \vdots \\ a_{n-r}'y \end{bmatrix} = A'y.$$

- If $\boldsymbol{a}'\boldsymbol{X} = \mathbf{0}'$, then $\boldsymbol{a}'\boldsymbol{y}$ is known as an error contrast.

- Thus, $w_1, \ldots, w_{n-r}$ comprise a set of $n - r$ error contrasts.

- Because $(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \mathbf{0}$, the elements of $(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ are each error contrasts.

- Because $rank(\boldsymbol{I} - \boldsymbol{P_X}) = n - rank(\boldsymbol{X}) = n - r$, there exists a set of $n - r$ linearly independent rows of $\boldsymbol{I} - \boldsymbol{P_X}$ that can be used in step 1 of the REML method to get $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}$.

- If we do use a subset of rows of $\boldsymbol{I} - \boldsymbol{P_X}$ to get $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}$, the error contrasts $w_1 = \boldsymbol{a}_1'\boldsymbol{y}, \ldots, w_{n-r} = \boldsymbol{a}_{n-r}'\boldsymbol{y}$ will be a subset of the elements of $(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} = \boldsymbol{y} - \hat{\boldsymbol{y}}$, the residual vector.

- This is why it makes sense to call the procedure Residual Maximum Likelihood.

- Note that

$$
\begin{aligned}
\boldsymbol{w} &= \boldsymbol{A}'\boldsymbol{y} \\
&= \boldsymbol{A}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{\varepsilon} \\
&= \boldsymbol{0} + \boldsymbol{A}'\boldsymbol{\varepsilon} \\
&= \boldsymbol{A}'\boldsymbol{\varepsilon}.
\end{aligned}
$$

- Thus, $\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{\varepsilon} \sim N(\boldsymbol{A}'\boldsymbol{0}, \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}) \stackrel{d}{=} N(\boldsymbol{0}, \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})$, and the distribution of $\boldsymbol{w}$ depends on $\boldsymbol{\theta}$ but not $\boldsymbol{\beta}$.

The log likelihood function in this case is

$$l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w}) = -\frac{n-r}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}| - \frac{1}{2}\boldsymbol{w}'(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{w}.$$

An MLE for $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, can be found in the general case using numerical methods to obtain the REML estimate of $\boldsymbol{\theta}$.

Let

$$\hat{\boldsymbol{\theta}} = \arg\max\{l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{y}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}.$$

Once a REML estimate of $\boldsymbol{\theta}$ (and thus $\boldsymbol{\Sigma}$) has been obtained, the BLUE of an estimable $\boldsymbol{C}\boldsymbol{\beta}$ if $\boldsymbol{\Sigma}$ were unknown can be approximated by

$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_{\text{REML}} = \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y},$$

where $\hat{\boldsymbol{\Sigma}}$ is $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\theta}}$ (the REML estimate of $\boldsymbol{\theta}$) in place of $\boldsymbol{\theta}$.

Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are each $n \times (n-r)$ matrices satisfying

$$rank(\boldsymbol{A}) = rank(\boldsymbol{B}) = n - r \quad \text{and} \quad \boldsymbol{A}'\boldsymbol{X} = \boldsymbol{B}'\boldsymbol{X} = \boldsymbol{0}.$$

Let

$$\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{y} \quad \text{and} \quad \boldsymbol{v} = \boldsymbol{B}'\boldsymbol{y}.$$

Prove that

$$\hat{\boldsymbol{\theta}} \text{ maximizes } l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w}) \text{ over } \boldsymbol{\theta} \in \boldsymbol{\Theta}$$

iff

$$\hat{\boldsymbol{\theta}} \text{ maximizes } l_{\boldsymbol{v}}(\boldsymbol{\theta}|\boldsymbol{v}) \text{ over } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Proof:

$\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{0} \Longrightarrow \boldsymbol{X}'\boldsymbol{A} = \boldsymbol{0} \Longrightarrow$ each column of $\boldsymbol{A} \in \mathcal{N}(\boldsymbol{X}')$.

$dim(\mathcal{N}(\boldsymbol{X}')) = n - rank(\boldsymbol{X}) = n - r.$

$\therefore$ the $n - r$ LI columns of $\boldsymbol{A}$ form a basis for $\mathcal{N}(\boldsymbol{X}')$.

$\boldsymbol{B}'\boldsymbol{X} = \boldsymbol{0} \Longrightarrow \boldsymbol{X}'\boldsymbol{B} = \boldsymbol{0} \Longrightarrow$ columns of $\boldsymbol{B} \in \mathcal{N}(\boldsymbol{X}')$.

$\therefore \exists$ an $(n - r) \times (n - r)$ matrix $\boldsymbol{M} \ni \boldsymbol{A}\boldsymbol{M} = \boldsymbol{B}$.

Note that $\boldsymbol{M}$, an $(n-r) \times (n-r)$ matrix, is nonsingular $\because$

$$
\begin{aligned}
n - r &= rank(\boldsymbol{B}) = rank(\boldsymbol{AM}) \\
&\leq rank(\boldsymbol{M}) \leq n - r \\
\Longrightarrow \quad & \quad rank(\boldsymbol{M}) = n - r. \\
\therefore \boldsymbol{v}'(\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B})^{-1}\boldsymbol{v} &= \boldsymbol{y}'\boldsymbol{B}(\boldsymbol{M}'\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{AM})^{-1}\boldsymbol{B}'\boldsymbol{y} \\
&= \boldsymbol{y}'\boldsymbol{AMM}^{-1}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}(\boldsymbol{M}')^{-1}\boldsymbol{M}'\boldsymbol{A}'\boldsymbol{y} \\
&= \boldsymbol{y}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{y} = \boldsymbol{w}'(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{w}.
\end{aligned}
$$

Also

$$
\begin{aligned}
|\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B}| &= |\boldsymbol{M}'\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{M}| \\
&= |\boldsymbol{M}'||\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}||\boldsymbol{M}| \\
&= |\boldsymbol{M}|^2|\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}|.
\end{aligned}
$$

Now note that

$$
\begin{aligned}
l_{\boldsymbol{v}}(\boldsymbol{\theta}|\boldsymbol{v}) &= -\frac{n-r}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B}| - \frac{1}{2}\boldsymbol{v}'(\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B})^{-1}\boldsymbol{v} \\
&= -\frac{n-r}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{M}|^2|\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}|) - \frac{1}{2}\boldsymbol{w}'(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{w} \\
&= -\frac{1}{2}\log(|\boldsymbol{M}|^2) - \frac{n-r}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}|) \\
&\qquad - \frac{1}{2}\boldsymbol{w}'(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{w} \\
&= -\log|\boldsymbol{M}| + l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w}).
\end{aligned}
$$

Because $\boldsymbol{M}$ is free of $\boldsymbol{\theta}$, the result follows. □

Consider the special case

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

Prove that the REML estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}}{n - r}.$$

Proof:

$\boldsymbol{I} - \boldsymbol{P_X}$ is symmetric. Thus, by the Spectral Decomposition Theorem,

$$\boldsymbol{I} - \boldsymbol{P_X} = \sum_{j=1}^{n} \lambda_j \boldsymbol{q}_j \boldsymbol{q}_j'$$

where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of $\boldsymbol{I} - \boldsymbol{P_X}$ and $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are orthonormal eigenvectors.

Because $\boldsymbol{I} - \boldsymbol{P_X}$ is idempotent and of rank $n - r$, $n - r$ of the λ_j values equal 1 and the other r equal 0.

Define $\boldsymbol{Q}$ to be the matrix whose columns are the eigenvectors corresponding to the $n - r$ nonzero eigenvalues.

Then $\boldsymbol{Q}$ is $n \times (n-r)$, $\boldsymbol{QQ}' = \boldsymbol{I} - \boldsymbol{P_X}$, and $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$.

Thus,

$$rank(\boldsymbol{Q}) = n - r \quad \text{and} \quad \boldsymbol{QQ}'\boldsymbol{X} = (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{P_X X} = \boldsymbol{0}.$$

Multiplying on the left by $\boldsymbol{Q}'$

$$\begin{aligned} &\Longrightarrow \boldsymbol{Q}'\boldsymbol{QQ}'\boldsymbol{X} = \boldsymbol{Q}'\boldsymbol{0} \\ &\Longrightarrow \boldsymbol{IQ}'\boldsymbol{X} = \boldsymbol{0} \\ &\Longrightarrow \boldsymbol{Q}'\boldsymbol{X} = \boldsymbol{0}. \end{aligned}$$

$\therefore$ the REML estimator of σ^2 can be obtained by finding the maximizer of the likelihood based on the data

$$\begin{aligned} \boldsymbol{w} \equiv \boldsymbol{Q}'\boldsymbol{y} &\sim N(\boldsymbol{Q}'\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{Q}'\sigma^2\boldsymbol{I}\boldsymbol{Q}) \\ &\stackrel{d}{=} N(\boldsymbol{0}, \sigma^2\boldsymbol{I}). \end{aligned}$$

$$
\begin{aligned}
l_{\boldsymbol{w}}(\sigma^2|\boldsymbol{w}) &= -\frac{n-r}{2}\log(2\pi) - \frac{n-r}{2}\log(\sigma^2) - \frac{1}{2}\boldsymbol{w}'\boldsymbol{w}/\sigma^2 \\
\frac{\partial l_{\boldsymbol{w}}(\sigma^2|\boldsymbol{w})}{\partial \sigma^2} &= -\frac{n-r}{2\sigma^2} + \frac{\boldsymbol{w}'\boldsymbol{w}}{2\sigma^4} \\
\left.\frac{\partial l_{\boldsymbol{w}}(\sigma^2|\boldsymbol{w})}{\partial \sigma^2}\right|_{\sigma^2=\hat{\sigma}^2} &= 0 \Longleftrightarrow \hat{\sigma}^2 = \frac{\boldsymbol{w}'\boldsymbol{w}}{n-r}
\end{aligned}
$$

$\therefore$ the MLE of σ^2 based on $\boldsymbol{w}$ and the REML estimator of σ^2 based on $\boldsymbol{y}$ is

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\boldsymbol{w}'\boldsymbol{w}}{n-r} = \frac{(\boldsymbol{Q}'\boldsymbol{y})'(\boldsymbol{Q}'\boldsymbol{y})}{n-r} \\ &= \frac{\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{y}}{n-r} = \frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}}{n-r}.\end{aligned}$$

□

REML Theorem:

Suppose $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ a PD, symmetric matrix whose entries are known functions of an unknown parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Furthermore, suppose $rank(\boldsymbol{X}) = r$ and $\tilde{\boldsymbol{X}}$ is any $n \times r$ matrix consisting of any set of r LI columns of $\boldsymbol{X}$.

REML Theorem:

Let A be any $n \times (n-r)$ matrix satisfying

$$rank(\boldsymbol{A}) = n - r \quad \text{and} \quad \boldsymbol{A}'\boldsymbol{X} = \boldsymbol{0}.$$

Let

$$\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}).$$

Then $\hat{\boldsymbol{\theta}}$ maximizes $l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w})$ over $\boldsymbol{\theta} \in \boldsymbol{\Theta} \Longleftrightarrow \hat{\boldsymbol{\theta}}$ maximizes, over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$g(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})),$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$.

We have previously shown:

1. $\exists$ an $n \times (n-r)$ matrix $\underset{n \times m}{\boldsymbol{Q}} \ni \boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}, \boldsymbol{Q}'\boldsymbol{X} = \boldsymbol{0}$, and $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$, and

2. Any $n \times (n-r)$ matrix $\boldsymbol{A}$ of rank $n-r$ satisfying $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{0}$ leads to the same REML estimator $\hat{\boldsymbol{\theta}}$.

Thus, without loss of generality, we may assume the matrix $\boldsymbol{A}$ in the REML Theorem is $\boldsymbol{Q}$.

We will now prove 3 lemmas and then use those lemmas to prove the REML Theorem.

Lemma R1:

Let $\boldsymbol{G} = \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}$ and suppose $\boldsymbol{A}$ is and $n \times (n-r)$ matrix satisfying

$$\boldsymbol{A}'\boldsymbol{A} = \underset{m \times m}{\boldsymbol{I}} \quad \text{and} \quad \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}.$$

Then

$$|[\boldsymbol{A}, \boldsymbol{G}]|^2 = |\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-1}.$$

Proof of Lemma R1:

$$\begin{aligned}
|[\boldsymbol{A},\boldsymbol{G}]|^2 &= |[\boldsymbol{A},\boldsymbol{G}]||[\boldsymbol{A},\boldsymbol{G}]| \\
&= |[\boldsymbol{A},\boldsymbol{G}]'||[\boldsymbol{A},\boldsymbol{G}]| \\
&= \begin{vmatrix} \boldsymbol{A}'\boldsymbol{A} & \boldsymbol{A}'\boldsymbol{G} \\ \boldsymbol{G}'\boldsymbol{A} & \boldsymbol{G}'\boldsymbol{G} \end{vmatrix} = \begin{vmatrix} \boldsymbol{I} & \boldsymbol{A}'\boldsymbol{G} \\ \boldsymbol{G}'\boldsymbol{A} & \boldsymbol{G}'\boldsymbol{G} \end{vmatrix} \\
&= |\boldsymbol{I}||\boldsymbol{G}'\boldsymbol{G} - \boldsymbol{G}'\boldsymbol{A}\boldsymbol{A}'\boldsymbol{G}|
\end{aligned}$$

(by our result on the determinant of a partitioned matrix)

$$
\begin{aligned}
&= |G'G - G'(I - P_X)G| \\
&= |G'P_XG| = |G'P_{\tilde{X}}G| \\
&= |(\tilde{X}'\Sigma\tilde{X})^{-1}\tilde{X}'\Sigma^{-1}\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\Sigma^{-1}\tilde{X}(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}| \\
&= |[(\tilde{X}'\Sigma\tilde{X})^{-1}(\tilde{X}'\Sigma^{-1}\tilde{X})](\tilde{X}'\tilde{X})^{-1}[(\tilde{X}'\Sigma^{-1}\tilde{X})(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}]| \\
&= |(\tilde{X}'\tilde{X})^{-1}| = |\tilde{X}'\tilde{X}|^{-1}.
\end{aligned}
$$

□

Lemma R2:

$$|\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}| = |\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}||\boldsymbol{\Sigma}||\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-1},$$

where $\boldsymbol{A}, \tilde{\boldsymbol{X}}$, and $\boldsymbol{\Sigma}$ are as previously defined.

Proof of Lemma R2:

Again define

$$\boldsymbol{G} = \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}.$$

Show that

(1) $\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{G} = \boldsymbol{0}$, and

(2) $\boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} = (\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}$.

$$
\begin{aligned}
\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{G} &= \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1} \\
&= \boldsymbol{A}'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1} \\
&= \boldsymbol{0} \because \boldsymbol{A}'\boldsymbol{X} = \boldsymbol{0} \Longrightarrow \boldsymbol{A}'\tilde{\boldsymbol{X}} = \boldsymbol{0}. \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} &= (\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1} \\
&= (\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1} \\
&= (\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}. \qquad (2)
\end{aligned}
$$

Next show that

$$|[\boldsymbol{A}, \boldsymbol{G}]|^2 |\boldsymbol{\Sigma}| = |\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}||\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}|^{-1}.$$

$$
\begin{aligned}
|[\boldsymbol{A},\boldsymbol{G}]|^2|\boldsymbol{\Sigma}| &= |[\boldsymbol{A},\boldsymbol{G}]'||\boldsymbol{\Sigma}||[\boldsymbol{A},\boldsymbol{G}]| \\
&= \left| \begin{bmatrix} \boldsymbol{A}' \\ \boldsymbol{G}' \end{bmatrix} \boldsymbol{\Sigma}[\boldsymbol{A},\boldsymbol{G}] \right| \\
&= \begin{vmatrix} \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{G} \\ \boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} \end{vmatrix} \\
&= \begin{vmatrix} \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & (\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1} \end{vmatrix} \text{ by (1) and (2)} \\
&= |\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}||\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}|^{-1}
\end{aligned}
$$

Now use Lemma R1 to complete the proof of Lemma R2.

We have

$$
\begin{aligned}
|[\boldsymbol{A}, \boldsymbol{G}]|^2 |\boldsymbol{\Sigma}| &= |\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}||\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}|^{-1} \\
\implies |\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}| &= |\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}||\boldsymbol{\Sigma}||[\boldsymbol{A}, \boldsymbol{G}]|^2 \\
&= |\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}||\boldsymbol{\Sigma}||\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-1}
\end{aligned}
$$

by Lemma R1.

□

Lemma R3:

$$A(A'\Sigma A)^{-1}A' = \Sigma^{-1} - \Sigma^{-1}\tilde{X}(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\tilde{X}'\Sigma^{-1},$$

where A, Σ, and $\tilde{X}$ are as defined previously.

Proof of Lemma R3:

$$
\begin{aligned}
[\boldsymbol{A},\boldsymbol{G}]^{-1}\boldsymbol{\Sigma}^{-1}([\boldsymbol{A},\boldsymbol{G}]')^{-1} &= ([\boldsymbol{A},\boldsymbol{G}]'\boldsymbol{\Sigma}[\boldsymbol{A},\boldsymbol{G}])^{-1} \\
&= \begin{bmatrix} \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{G} \\ \boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & (\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} (\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}} \end{bmatrix}.
\end{aligned}
$$

Now multiplying on the left by $[\boldsymbol{A}, \boldsymbol{G}]$ and on the right by $[\boldsymbol{A}, \boldsymbol{G}]'$ yields

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= [\boldsymbol{A}, \boldsymbol{G}] \begin{bmatrix} (\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}' \\ \boldsymbol{G}' \end{bmatrix} \\ &= \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{A}' + \boldsymbol{G}\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}\boldsymbol{G}'. \qquad (3) \end{aligned}$$

Now

$$
\begin{aligned}
G\tilde{X}'\Sigma^{-1}\tilde{X}G' &= \Sigma^{-1}\tilde{X}(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\tilde{X}'\Sigma^{-1}\tilde{X}(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\tilde{X}'\Sigma^{-1} \\
&= \Sigma^{-1}\tilde{X}(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\tilde{X}'\Sigma^{-1}. \qquad (4)
\end{aligned}
$$

Combining (3) and (4) yields

$$
A(A'\Sigma A)^{-1}A' = \Sigma^{-1} - \Sigma^{-1}\tilde{X}(\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\tilde{X}'\Sigma^{-1}.
$$

□

Now use Lemmas R2 and R3 to prove the REML Theorem.

$$
\begin{aligned}
l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w}) &= -\frac{n-r}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}| - \frac{1}{2}\boldsymbol{w}'(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{w} \\
&= \text{constant}_1 - \frac{1}{2}\log(|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}||\boldsymbol{\Sigma}||\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|^{-1}) \\
&\qquad -\frac{1}{2}\boldsymbol{y}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{y} \\
&= \text{constant}_2 - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}| \\
&\qquad -\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1})\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \text{constant}_2 - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}| \\
& \quad -\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{I} - \boldsymbol{\Sigma}^{-1/2}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1/2})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} \\
= \quad & \text{constant}_2 - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}| \\
& \quad -\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{\Sigma}^{-1/2}\tilde{\boldsymbol{X}}})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}.
\end{aligned}
$$

Now

$$\begin{aligned}
\mathcal{C}(\boldsymbol{\Sigma}^{-1/2}\tilde{\boldsymbol{X}}) &= \mathcal{C}(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}) \\
\therefore \boldsymbol{P}_{\boldsymbol{\Sigma}^{-1/2}\tilde{\boldsymbol{X}}} &= \boldsymbol{P}_{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}} \\
\therefore \boldsymbol{y}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{\Sigma}^{-1/2}\tilde{\boldsymbol{X}}})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} &= \boldsymbol{y}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} \\
&= \|(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\|^2
\end{aligned}$$

$$
\begin{aligned}
&= \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\|^2 \\
&= \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\|^2 \\
&= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \\
&= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})).
\end{aligned}
$$

$\therefore$ we have

$$
\begin{aligned}
l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w}) &= \text{constant}_2 - \frac{1}{2}\log|\boldsymbol{\Sigma}| \\
&\quad -\frac{1}{2}\log|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}| \\
&\quad -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \\
&= \text{constant}_2 + \mathrm{g}(\boldsymbol{\theta}).
\end{aligned}
$$

$\therefore$ $\hat{\boldsymbol{\theta}}$ maximizes $l_{\boldsymbol{w}}(\boldsymbol{\theta}|\boldsymbol{w})$ over $\boldsymbol{\theta} \in \boldsymbol{\Theta} \iff \hat{\boldsymbol{\theta}}$ maximizes $g(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

$\square$