

ML and REML Variance Component Estimation

Suppose

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

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 $\boldsymbol{\theta} \in \Theta$.

For example, consider

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

where $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ are known matrices,

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

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

The likelihood function is

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

and the log likelihood function is

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

Based on previous results,

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

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

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

Thus, the profile log likelihood for θ is

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

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

Let

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

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

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

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

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

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

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

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

It follows that

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

Note that

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

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

- We have seen by example that the MLE of the variance component vector θ can be biased.
- For example, for the case of $\Sigma = \sigma^2 \mathbf{I}$, where $\theta = [\sigma^2]$, the MLE of σ^2 is

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

This MLE for σ^2 is often criticized for “failing to account for the loss of degrees of freedom needed to estimate β .”

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

A familiar special case: Suppose

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

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 $n - \text{rank}(\mathbf{X}) = n - r$ linearly independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_{n-r}$ such that $\mathbf{a}'_i \mathbf{X} = \mathbf{0}'$ for all $i = 1, \dots, n - r$.
- Find the maximum likelihood estimate of θ using $w_1 \equiv \mathbf{a}'_1 \mathbf{y}, \dots, w_{n-r} \equiv \mathbf{a}'_{n-r} \mathbf{y}$ as data.

$$\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_{n-r}] \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_{n-r} \end{bmatrix} = \begin{bmatrix} \mathbf{a}'_1 \mathbf{y} \\ \vdots \\ \mathbf{a}'_{n-r} \mathbf{y} \end{bmatrix} = \mathbf{A}' \mathbf{y}.$$

- If $\mathbf{a}'\mathbf{X} = \mathbf{0}'$, then $\mathbf{a}'\mathbf{y}$ is known as an error contrast.
- Thus, w_1, \dots, w_{n-r} comprise a set of $n - r$ error contrasts.
- Because $(\mathbf{I} - \mathbf{P}_X)\mathbf{X} = \mathbf{X} - \mathbf{P}_X\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$, the elements of $(\mathbf{I} - \mathbf{P}_X)\mathbf{y} = \mathbf{y} - \mathbf{P}_X\mathbf{y} = \mathbf{y} - \hat{\mathbf{y}}$ are each error contrasts.

- Because $\text{rank}(\mathbf{I} - \mathbf{P}_X) = n - \text{rank}(\mathbf{X}) = n - r$, there exists a set of $n - r$ linearly independent rows of $\mathbf{I} - \mathbf{P}_X$ that can be used in step 1 of the REML method to get $\mathbf{a}_1, \dots, \mathbf{a}_{n-r}$.
- If we do use a subset of rows of $\mathbf{I} - \mathbf{P}_X$ to get $\mathbf{a}_1, \dots, \mathbf{a}_{n-r}$, the error contrasts $w_1 = \mathbf{a}'_1 \mathbf{y}, \dots, w_{n-r} = \mathbf{a}'_{n-r} \mathbf{y}$ will be a subset of the elements of $(\mathbf{I} - \mathbf{P}_X)\mathbf{y} = \mathbf{y} - \hat{\mathbf{y}}$, the residual vector.
- This is why it makes sense to call the procedure Residual Maximum Likelihood.

- Note that

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

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

The log likelihood function in this case is

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

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

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

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

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

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

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

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

Let

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

Prove that

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

iff

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

Proof:

$$\mathbf{A}'\mathbf{X} = \mathbf{0} \implies \mathbf{X}'\mathbf{A} = \mathbf{0} \implies \text{each column of } \mathbf{A} \in \mathcal{N}(\mathbf{X}').$$

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

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

$$\mathbf{B}'\mathbf{X} = \mathbf{0} \implies \mathbf{X}'\mathbf{B} = \mathbf{0} \implies \text{columns of } \mathbf{B} \in \mathcal{N}(\mathbf{X}').$$

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

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

$$n - r = \text{rank}(\mathbf{B}) = \text{rank}(\mathbf{AM})$$

$$\leq \text{rank}(\mathbf{M}) \leq n - r$$

$$\implies \text{rank}(\mathbf{M}) = n - r.$$

$$\begin{aligned} \therefore \mathbf{v}'(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{v} &= \mathbf{y}'\mathbf{B}(\mathbf{M}'\mathbf{A}'\Sigma\mathbf{AM})^{-1}\mathbf{B}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{A}\mathbf{M}\mathbf{M}^{-1}(\mathbf{A}'\Sigma\mathbf{A})^{-1}(\mathbf{M}')^{-1}\mathbf{M}'\mathbf{A}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{A}(\mathbf{A}'\Sigma\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \mathbf{w}'(\mathbf{A}'\Sigma\mathbf{A})^{-1}\mathbf{w}. \end{aligned}$$

Also

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

Now note that

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

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

Consider the special case

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

Prove that the REML estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}}{n - r}.$$

REML Theorem:

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

REML Theorem:

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

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

Let

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

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

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

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

We have previously shown:

1. \exists an $n \times (n - r)$ matrix $\mathbf{Q} \underset{n \times m}{\ni} \mathbf{Q}\mathbf{Q}' = \mathbf{I} - \mathbf{P}_X, \mathbf{Q}'\mathbf{X} = \mathbf{0}$, and $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$,
and
2. Any $n \times (n - r)$ matrix \mathbf{A} of rank $n - r$ satisfying $\mathbf{A}'\mathbf{X} = \mathbf{0}$ leads to the same REML estimator $\hat{\boldsymbol{\theta}}$.

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

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

Lemma R1:

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

$$\mathbf{A}'\mathbf{A} = \mathbf{I}_{m \times m} \quad \text{and} \quad \mathbf{A}\mathbf{A}' = \mathbf{I} - \mathbf{P}_X.$$

Then

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

Proof of Lemma R1:

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

(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:

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

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

Proof of Lemma R2:

Again define

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

Show that

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

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

Next show that

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

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

Lemma R3:

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

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

Proof of Lemma R3:

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

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

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

Now

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

Combining (3) and (4) yields

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

□

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

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

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

Now

$$\mathcal{C}(\Sigma^{-1/2}\tilde{X}) = \mathcal{C}(\Sigma^{-1/2}X)$$

$$\therefore P_{\Sigma^{-1/2}\tilde{X}} = P_{\Sigma^{-1/2}X}$$

$$\begin{aligned}\therefore \mathbf{y}'\Sigma^{-1/2}(\mathbf{I} - P_{\Sigma^{-1/2}\tilde{X}})\Sigma^{-1/2}\mathbf{y} &= \mathbf{y}'\Sigma^{-1/2}(\mathbf{I} - P_{\Sigma^{-1/2}X})\Sigma^{-1/2}\mathbf{y} \\ &= \|(\mathbf{I} - P_{\Sigma^{-1/2}X})\Sigma^{-1/2}\mathbf{y}\|^2\end{aligned}$$

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

∴ we have

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

∴ $\hat{\boldsymbol{\theta}}$ maximizes $l_w(\boldsymbol{\theta}|\mathbf{w})$ over $\boldsymbol{\theta} \in \Theta \iff \hat{\boldsymbol{\theta}}$ maximizes $g(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \Theta$.

□