The simple model:

$$Y = X\gamma + H\beta + e$$

where

$Y : N \times 1,$

$X : N \times p,$

$\gamma : p \times 1, \gamma$ is the fixed effect,

$H : N \times q,$

$\beta : q \times 1, \beta \sim MN(0, \tau R),$

and $e : N \times 1, e \sim MN(0, \sigma I).$

Let $U = A^T Y$ with the condition that $A^T A = I_{N-p}$ and $AA^T = I - P_X = I - X(X^T X)^{-1} X^T$, we

have

$$E(U) = E(A^T Y)$$
$$= E(A^T X\gamma + A^T H\beta + A^T e)$$
$$= A^T X\gamma$$

Consider that,

$$(A^T X\gamma)^T (A^T X\gamma) = \gamma^T X^T A A^T X\gamma$$
$$= \gamma^T X^T \underbrace{(I - P_X)X}_{0} \gamma$$
$$= 0$$

Therefore $A^T X\gamma = 0$, so $E(U) = 0$.

Define $\mathrm{Var}(Y) = V = \tau HRH^T + \sigma I$, so we have $\mathrm{Var}(U) = A^T V A$.

Since U and β are normal distributed, the joint distribution of them is also normal distributed,

the covariance of U and β is

$$Cov(U, \beta) = Cov(A^T X\gamma + A^T H\beta + A^T e, \beta)$$
$$= Cov(A^T H\beta, \beta)$$
$$= A^T H Var(\beta) = \tau A^T H R$$

Therefore, the joint distribution of $(U, \beta)^T$ is

$$\begin{pmatrix} U \\ \beta \end{pmatrix} \sim MN\left(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} A^T V A & \tau A^T H R \\ \tau R H^T A & \tau R \end{pmatrix}\right)$$

According to $E(\mu_1|\mu_2) = \mu_1 + \Sigma_{12}(\Sigma_{22})^{-1}\mu_2$ and $Var(\mu_1|\mu_2) = \Sigma_{11}(\Sigma_{22})^{-1}\Sigma_{21}$.

$$E(U|\beta) = 0 + \tau A^T H R \tau^{-1} R^{-1} \beta$$

$$= A^T H \beta$$

$$Var(U|\beta) = A^T V A - \tau A^T H R \tau^{-1} R^{-1} \tau R H^T A$$

$$= A^T V A - A^T (\tau H R H^T) A$$

$$= A^T (V - \tau H R H^T) A$$

$$= A^T \cdot \sigma I \cdot A$$

$$= \sigma I$$

Therefore, we have the conditional distribution of $U|\beta$ which is

$$U|\beta \sim MN(A^T H \beta, \sigma I_{N-p})$$

For $\beta|U$, we have

$$E(\beta|U) = 0 + \tau R H^T A \cdot (A^T V A)^{-1} \cdot A^T Y$$

$$= \tau R H^T \cdot A(A^T V A)^{-1} A^T \cdot Y$$

$$Var(\beta|U) = \tau R - \tau R H^T A \cdot (A^T V A)^{-1} \cdot \tau A^T H R$$

$$= \tau R - \tau^2 R H^T \cdot A(A^T V A)^{-1} A^T \cdot \tau H R$$

To show that $A(A^T V A)^{-1} A^T = P = V^{-1} - V^{-1}X(X^T V^{-1}X)^{-1}X^T V^{-1}$, we constructing the $n \times p$ matrix $\mathbf{G} = \mathbf{V^{-1}X(X^T V^{-1}X)^{-1}}$ and the $n \times n$ matrix $\mathbf{B} = [\mathbf{A}|\mathbf{G}]$. First the product,

$$\mathbf{B}^T\mathbf{B} = \begin{bmatrix} \mathbf{A}^T \\ \mathbf{G}^T \end{bmatrix} [\mathbf{A} \quad \mathbf{G}] = \begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{G} \\ \mathbf{G}^T\mathbf{A} & \mathbf{G}^T\mathbf{G} \end{bmatrix}$$

then the determinant can be rewritten as

$$|\mathbf{B}|^2 = |\mathbf{B'B}| = |\mathbf{A'A}| \times |\mathbf{G'G} - \mathbf{G'A(A'A)}^{-1}\mathbf{A'G}|$$

$$= |\mathbf{I}| \times |\mathbf{G'G} - \mathbf{G'(I - P_X)G}| = |\mathbf{G'}P_{\mathbf{X}}\mathbf{G}|$$

$$= |(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}\mathbf{X(X'X)}^{-1}\mathbf{X'V}^{-1}\mathbf{X(X'V}^{-1}\mathbf{X})^{-1}| = |(\mathbf{X'X})^{-1}|.$$

$$(A^T V G)^T (A^T V G) = G^T V A A^T V G$$
$$= G^T V \underbrace{(I - P_X) V V^{-1} X}_{0} (X^T V^{-1} X)^{-1}$$
$$= 0$$

which also says that **B** is nonsingular.

$$V^{-1} = B(B^T V B)^{-1} B^T$$

$$= [A \quad G] \begin{bmatrix} A^T V A & 0 \\ 0 & G^T V G \end{bmatrix} \begin{bmatrix} A^T \\ G^T \end{bmatrix}$$

$$= A(A^T V A)^{-1} A^T + G(G^T V G)^{-1} G^T$$

$$= A(A^T V A) - 1 A^T + G \underbrace{((X^T V^{-1} X)^{-1} X^T V^{-1} V V^{-1} X (X^T V^{-1} X)^{-1})^{-1}}_{X^T V^{-1} X} G^T$$

$$= A(A^T V A) - 1 A^T + V^{-1} X (X^T V^{-1} X)^{-1} \cdot X^T V^{-1} X \cdot (X^T V^{-1} X)^{-1} X^T V^{-1}$$

$$= A(A^T V A) - 1 A^T + V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}$$

so we have $A(A^T V A)^{-1} A^T = P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}$. Therefore, the expectation and variance of $\beta|U$ can be simplified as

$$E(\beta|U) = \tau R H^T P_X Y$$

$$Var(\beta|U) = \tau R - \tau^2 R H^T P_X H R$$

In another words, we have,

$$\beta|U \sim MN(\tau R H^T P Y, \tau R - \tau^2 R H^T P H R))$$

Define $\theta = \{\sigma, \tau\}$, according to the EM algorithm, we need to first compute the expectation of $\log \quad L(\theta^{(t)}|U, \beta)$ given U and $\theta^{(t-1)}$,

$$\log L(\theta^{(t)}|U, \beta) = \log f(U, \beta|\theta^{(t)})$$
$$= \log\{f(U|\beta, \theta^{(t)}) f(\beta|\theta^{(t)})\}$$
$$= \log f(U|\beta, \theta^{(t)}) + \log f(\beta|\theta^{(t)})$$

1. Estimate τ Take the expectation of $\log L(\theta^{(t)}|U, \beta)$ with $\beta|U, \theta^{(t-1)}$, we have

$$E_\beta\left[\log L(\theta^{(t)}|U,\beta)\Big|U,\theta^{(t-1)}\right] = E_\beta\left[\log f(U|\beta,\theta^{(t)})\Big|U,\theta^{(t-1)}\right] - \frac{1}{2}q\log\tau - \frac{1}{2}\log|R| - \frac{1}{2\tau}E_\beta\left[(\beta^T R^{-1}\beta)\Big|U,\theta^{(t-1)}\right]$$

3

Since the $E(\varepsilon^T \Lambda \varepsilon) = E(\varepsilon)^T \Lambda E(\varepsilon) + tr(\Lambda Var(\varepsilon))$, (1) can be simplified as

$$E_{\beta|U,\theta^{(t-1)}}(\log L(\theta^{(t)}|U,\beta)) = E_\beta \left[\log f(U|\beta,\theta^{(t)}) \Big| U, \theta^{(t-1)} \right] - \frac{1}{2} q \log \tau - \frac{1}{2} \log |R|$$
$$- \frac{1}{2\tau} \left[E(\beta|U,\theta^{(t-1)})^T R^{-1} E(\beta|U,\theta^{(t-1)}) + tr(R^{-1} Var(\beta|U,\theta^{(t-1)})) \right]$$

To maximize the expectation above, we let

$$\frac{\partial E_{\beta|U,\theta^{(t-1)}}(\log L(\theta^{(t)}|U,\beta))}{\partial \theta} = 0$$

It is easily to derive that

$$\tau^{(t)} = \frac{1}{q} \left[E(\beta|U,\theta^{(t-1)})^T R^{-1} E(\beta|U,\theta^{(t-1)}) + tr(R^{-1} Var(\beta|U,\theta^{(t-1)})) \right] \tag{1}$$

2. Estimate σ We can also get the estimation of σ by Define

$$Q = E_\beta \left[(\log L(\theta^{(t)}|U,\beta)) \Big| U, \theta^{(t-1)} \right]$$
$$= E_\beta \left[-\frac{1}{2} \left[(N-p) \log \sigma + 1 + \frac{1}{\sigma}(U - A^T H\beta)^T (U - A^T H\beta) \Big| U, \theta^{(t-1)} \right] \right.$$
$$+ E_\beta \left[\log f(\beta|\theta^{(t)}) \Big| U, \theta^{(t-1)} \right]$$

If we take the expectation on β, then

$$E_\beta[\log(L)|U,\theta^{(t-1)}]$$
$$= E_\beta \left[-\frac{1}{2}(N-p)\log \sigma + 1 + \frac{1}{\sigma}\left((Y-H\beta)^T AA^T (Y-H\beta) \right) + \log f(\beta|\theta^{(t)}) \Big| U, \theta^{(t-1)} \Big| U, \theta^{(t-1)} \right]$$
$$= -\frac{1}{2}(N-p)\log \sigma + 1 + \frac{1}{\sigma}\left((Y-H\tilde{\beta})^T (I - P_X)(Y-H\tilde{\beta}) + tr(H^T(I-P_X)H Var(\beta|U)) \right)$$
$$+ E_\beta \left[\log f(\beta|\theta^{(t)}) \Big| U, \theta^{(t-1)} \right] + 1$$

Let $\dfrac{\partial Q}{\partial \sigma} = 0$, we have

$$\sigma^{(t)} = \frac{1}{N-p} \left[(Y-H\tilde{\beta})^T (I-P_X)(Y-H\tilde{\beta}) + tr(H^T(I-P_X)H Var(\beta|U)) \right] \tag{2}$$

If the model is

$$Y = X\gamma + H_A \beta_A + H_B \beta_B + e$$

where

$Y : N \times 1,$

$X : N \times p,$

$\gamma : p \times 1$, γ is the fixed effect,

$H_A : N \times q_A$,

$\beta_A : q_A \times 1$, $\beta \sim MN(0, \tau_A R_A)$,

$H_B : N \times q_B$,

$\beta_B : q_B \times 1$, $\beta \sim MN(0, \tau_B R_B)$,

construct $H = [H_A H_B]$, $\beta = (\beta_A, \beta_B)^T$ and $R = \begin{bmatrix} \tau_A R_A & 0 \\ 0 & \tau_B R_B \end{bmatrix}$. According to the simple model,

we can easily get that

$$\begin{pmatrix} U \\ \beta \end{pmatrix} \sim MN\left(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} A^T V A & A^T H R \\ R H^T A & \tau R \end{pmatrix}\right)$$

and

$$E(U|\beta) = A^T H \beta$$

$$Var(U|\beta) = \sigma I_{N-p}$$

$$E(\beta|U) = R H^T P Y$$

$$Var(\beta|U) = R - R H^T P H R$$

The likelihood function is

$$\begin{aligned}
\log L(\theta^{(t)}|U, \beta) &= \log f(U, \beta|\theta^{(t)}) \\
&= \log f(U|\beta, \theta^{(t)}) + \log f(\beta|\theta^{(t)}) \\
&= \left(-\frac{1}{2}\right)\left[\log\left|\begin{bmatrix} \tau_A R_A & 0 \\ 0 & \tau_B R_B \end{bmatrix}\right| + (\beta_A \quad \beta_B)^T \begin{bmatrix} (\tau_A R_A)^{-1} & 0 \\ 0 & (\tau_B R_B)^{-1} \end{bmatrix} \begin{bmatrix} \beta_A \\ \beta_B \end{bmatrix}\right] \\
&\quad + \log f(U|\beta, \theta^{(t)}) \\
&= \log f(U|\beta, \theta^{(t)}) + \log f(\beta_A|\theta^{(t)}) + \log f(\beta_B|\theta^{(t)})
\end{aligned}$$

so the estimation of θ is

$$\tau_A^{(t)} = \frac{1}{q_A}\left[E(\beta_A|U, \theta^{(t-1)})^T R_A^{-1} E(\beta_A|U, \theta^{(t-1)}) + tr(R_A^{-1} Var(\beta_A|U, \theta^{(t-1)}))\right] \tag{3}$$

$$\tau_B^{(t)} = \frac{1}{q_B}\left[E(\beta_B|U, \theta^{(t-1)})^T R_B^{-1} E(\beta_B|U, \theta^{(t-1)}) + tr(R_B^{-1} Var(\beta_B|U, \theta^{(t-1)}))\right] \tag{4}$$

$$\sigma^{(t)} = \frac{1}{N-p}\left[(Y - H\tilde{\beta})^T (I - P_X)(Y - H\tilde{\beta}) + tr(H^T (I - P_X) H Var(\beta|U))\right] \tag{5}$$

APPENDIX

The variables we used in the algorithm are

$$P_X = X(X^T X)^{-1} X^T$$

$$V = \tau H R H^T + \sigma I$$

$$P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}$$

$$E(\beta|U) = \tau R H^T P Y$$

$$Var(\beta|U) = \tau R - \tau^2 R H^T P H R$$

$$\tau^{(t)} = \frac{1}{q} \left[E(\beta|U, \theta^{(t-1)})^T R^{-1} E(\beta|U, \theta^{(t-1)}) + tr(R^{-1} Var(\beta|U, \theta^{(t-1)})) \right]$$

$$\sigma^{(t)} = \frac{1}{N - p} \left[(Y - H\tilde{\beta})^T (I - P_X)(Y - H\tilde{\beta}) + tr(H^T (I - P_X) H Var(\beta|U)) \right]$$

Also, we have a algorithm for computing the inverse of a special form of matrix

$$(I + BC)^{-1} = I - B(I + CB)^{-1} C$$

So if we have to big matrix in the form of $(I + A)$ where A is $m \times m$, we can try to split A into $A = BC$ where B is $m \times n$ and n is much smaller than m.