The Derivation of EM alogrithm

The model is

$$Y = X\gamma + G_A + G_B + e$$

Y: $N \times 1$ matrix, records trait value.

X: $N \times p$ matrix, records the covariant.

γ: $p \times 1$, the effect of covariant.

G_A: the gene effect for Gene A, treated as an random effect. $G_A \sim N(0, \tau_A S_A)$, S_A is the similarity matrix which records the genetic similarity between individuals.

G_B: the gene effect for Gene B, also treated as an random effect. G_A and G_B are assumed to be independent.

e: $N \times 1$ matrix, the error term. $e \sim N(0, \sigma I)$.

Define $U = A^T Y$ with the restriction that $A^T A = I_{N-p}$ and $AA^T = I - P_X$.

It is easy to find out that : $E(U) = 0$ and $Var(U) = A^T V A$ where $V = Var(Y) = \tau_A S_A + \tau_B S_B + \sigma I$.

$$\begin{aligned}
Cov(U, G_A) =& Cov(A^T X\gamma + A^T G_A + A^T G_B + A^T e, G_A) \\
=& Cov(A^T G_A, G_A) \\
=& A^T Cov(G_A, G_A) \\
=& \tau_A A^T S_A
\end{aligned}$$

In the same way, we have$Cov(U, G_B) = \tau_B A^T S_B$ and $Cov(G_A, G_B) = 0$ since they are independent.

Therefore, the joint distribution of $(U, G_A, G_B)^T$ is

$$\begin{pmatrix} U \\ G_A \\ G_B \end{pmatrix} \sim MN\left(\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} A^T V A & \tau_A A^T S_A & \tau_B A^T S_B \\ \tau_A S_A A & \tau_A S_A & 0 \\ \tau_B S_B A & 0 & \tau_B S_B \end{pmatrix}\right)$$

so we can have the following conditional mean and variance,

1. the conditional mean and variance for U are

$$\begin{aligned}
E(U|G_A, G_B) =& A^T(G_A + G_B) \\
Var(U|G_A, G_B) =& \sigma I_{N-p}
\end{aligned}$$

2. the conditional mean and variance for G_A, since $Cov(G_A, G_B) = 0$

$$\begin{aligned}
E(G_A|G_B, U) =& E(G_A|U) \\
=& \tau_A S_A A(A^T V A)^{-1} A^T Y \\
Var(G_A|G_B, U) =& Var(G_A|U) \\
=& \tau_A S_A - \tau_A^2 S_A A(A^T V A)^{-1} A^T S_A
\end{aligned}$$

Simple algebra shows that $A(A^T V A)^{-1} A^T = P = V^{-1} - V^{-1} X(X^T V^{-1} X)^{-1} X^T V^{-1}$, so that

$$\begin{aligned}
E(G_A|G_B, U) =& \tau_A S_A P Y = g_A \\
Var(G_A|G_B, U) =& \tau_A S_A - \tau_A^2 S_A P S_A = v_A
\end{aligned}$$

3. Similarly, the conditional mean and variance for G_B are

$$\begin{aligned} E(G_B|G_A, U) =& \tau_B S_B P Y = g_B \\ Var(G_B|G_A, U) =& \tau_B S_B - \tau_B^2 S_B P S_B = v_B \end{aligned}$$

Define $\theta = \{\sigma, \tau_A, \tau_B\}$, according to the EM algorithm, we need to first compute the $\log L(\theta^{(t)}|U, G_A, G_B)$,

$$\begin{aligned} \log L(\theta^{(t)}|U, G_A, G_B) =& f(U, G_A, G_B|\theta^{(t)}) \\ =& \log f(U|G_A, G_B, \theta^{(t)}) + \log f(G_A|\theta^{(t)}) + \log f(G_B|\theta^{(t)}) \end{aligned}$$

Since S_A and S_B are singular, we have,

$$f(G_A) = \frac{1}{((2\pi)^{rank(S_A)}|\tau_A S_A|_+)^{\frac{1}{2}}} \exp\Big(-\frac{1}{2} G_A^T (\tau_A S_A)^- G_A \Big)$$

where $|\tau_A S_A|_+$ is the Pseudo-Determinant, and $(\tau_A S_A)^-)$ is the Generalized inverse.
Define $rank(S_A) = q_A$, $rank(S_B) = q_B$, we have,

$$\begin{aligned} f(G_A) =& \frac{1}{((2\pi)^{q_A}\tau_A^{q_A}|S_A|_+)^{\frac{1}{2}}} \exp\Big(-\frac{1}{2\tau_A} G_A^T (S_A)^- G_A \Big) \\ f(G_B) =& \frac{1}{((2\pi)^{q_B}\tau_B^{q_B}|S_B|_+)^{\frac{1}{2}}} \exp\Big(-\frac{1}{2\tau_B} G_B^T (S_B)^- G_B \Big) \end{aligned}$$

therefore,

$$\begin{aligned} \log f(G_A) =& constant - \frac{q_A}{2}\log\tau_A - \frac{1}{2}\log(|S_A|_+) - \frac{1}{2\tau_A} G_A^T S_A^- G_A \\ \log f(G_B) =& constant - \frac{q_B}{2}\log\tau_B - \frac{1}{2}\log(|S_B|_+) - \frac{1}{2\tau_B} G_B^T S_B^- G_B \\ \log f(U|G_A, G_B) =& constant - \frac{N-p}{2}\log\sigma - \frac{1}{2\sigma}\Big[(Y - G_A - G_B)^T (I - P_X)(Y - G_A - G_B)\Big] \end{aligned}$$

EM algorithm treats G_A and G_B as missing values. So instead of estimating G_A and G_B and then plugging them into the $\log L$, EM algorithm calculate the expectation of $\log L$ given U and $\theta^{(t-1)}$, then based on $E(\log L|U, \theta^{(t-1)})$, $\theta^{(t)}$ are calculated by taking partial derivative.

$$E\Big(\log L\Big|U, \theta^{(t-1)}\Big) = E\Big(\log f(G_A)\Big|U, \theta^{(t-1)}\Big) + E\Big(\log f(G_B)\Big|U, \theta^{(t-1)}\Big) + E\Big(\log f(U|G_A, G_B)\Big|\theta^{(t-1)}\Big)$$

It is easy to find out that to estimate $\tau_A^{(t)}$, we just need to consider $E\Big(\log f(G_A)\Big|U, \theta^{(t-1)}\Big)$, so let

$$\frac{\partial E\Big(\log f(G_A)\Big|U, \theta^{(t-1)}\Big)}{\partial \tau_A} = 0$$

we have,

$$-\frac{q_A}{2\tau_A} + \frac{1}{2\tau_A^2} E\Big(G_A^T S_A^- G_A\Big|U, \theta^{(t-1)}\Big) = 0$$

$$E\Big(G_A^T S_A^- G_A\Big|U, \theta^{(t-1)}\Big) = \big(g_A^{(t-1)}\big)^T S_A^- g_A^{(t-1)} + tr(S_A^- v_A^{(t-1)})$$

plugging the expression of $g_A^{(t-1)}$ and $v_A^{(t-1)}$, finally we have

$$\tau_A^{(t)} = \tau_A^{(t-1)} + \frac{[\tau_A^{t-1}]^2}{q_A}\Big[Y^T P S_A P Y - tr(S_A P)\Big]$$

In the same way, we can estimate $\tau_B^{(t)}$ by

$$\tau_B^{(t)} = \tau_B^{(t-1)} + \frac{[\tau_B^{t-1}]^2}{q_B}\Big[Y^T P S_B P Y - tr(S_B P)\Big]$$

To estimate $\sigma^{(t)}$, we just need to consider $E\Big(\log f(U|G_A, G_B)\Big|\theta^{(t-1)}\Big)$, so the final expression of $\sigma^{(t)}$ is

$$\sigma^{(t)} = \frac{1}{N-p}\Big\{\Big[(Y^*)^T(I-P_X)Y^*\Big] + tr\Big[(I-P_X)\Big(\tau_A^{(t-1)}S_A - (\tau_A^{(t-1)})^2 S_A P S_A + \tau_B^{(t-1)}S_B - (\tau_B^{(t-1)})^2 S_B P S_B\Big)\Big]\Big\}$$

where $Y^* = Y - \tau_A^{(t-1)}S_A P Y - \tau_B^{(t-1)}S_B P Y$