Multivariable Two-sample MR with tissue indicator

Without considering LD, the causal relationship is:

$$
\begin{aligned}
\widehat{\boldsymbol{\gamma}}_j \mid \boldsymbol{\gamma}_j, \widehat{\mathbf{S}}_{\gamma_j} &\sim \mathcal{N}\left(\boldsymbol{\gamma}_j, \widehat{\mathbf{S}}_{\gamma_j}\right) \\
\widehat{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}, \widehat{\mathbf{S}}_{\Gamma} &\sim \mathcal{N}\left(\boldsymbol{\Gamma}, \widehat{\mathbf{S}}_{\Gamma}\right)
\end{aligned} \tag{1}
$$

The causal relationship is:

$$
\Gamma_i = \alpha_i + \delta_1\beta_1\gamma_{i1} + \cdots + \delta_K\beta_K\gamma_{iK} \tag{2}
$$

Assume that

$$
\gamma_{ij} \sim \mathcal{N}(0, \sigma_\gamma^2),\ \ \delta_j \sim \text{Bernoulli}\left(\pi_j\right) \tag{3}
$$

The latent variable: $\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\delta}$ and the parameter set $\boldsymbol{\Theta} \stackrel{\text{def}}{=} \left\{\sigma_\alpha^2, \sigma_\gamma^2, \boldsymbol{\beta}, \boldsymbol{\pi}\right\}$. The complete-data likelihood of model is

$$
\begin{aligned}
&\log \Pr\left(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}_1, \cdots, \widehat{\boldsymbol{\gamma}}_K, \boldsymbol{\alpha} \mid \widehat{\mathbf{S}}_{\gamma_1}, \cdots, \widehat{\mathbf{S}}_{\gamma_K}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}; \boldsymbol{\Theta}\right) \\
=& \log \mathcal{N}\left(\widehat{\boldsymbol{\Gamma}} \mid \boldsymbol{\alpha} + \sum \delta_j\beta_j\boldsymbol{\gamma}_j, \widehat{\mathbf{S}}_{\Gamma}\right)\left(\prod_{j=1}^{K} \mathcal{N}\left(\widehat{\boldsymbol{\gamma}}_j \mid \boldsymbol{\gamma}_j, \widehat{\mathbf{S}}_{\gamma_j}\right)\right)\mathcal{N}\left(\boldsymbol{\alpha} \mid \mathbf{0}, \sigma_\alpha^2\mathbf{I_p}\right)\prod_{j=1}^{K}\mathcal{N}(\boldsymbol{\gamma}_j \mid 0, \sigma_\gamma^2\mathbf{I_p})\prod_{j=1}^{K}\mathbf{B}(1, \pi_j) \\
=& -\frac{p}{2}\log(2\pi) - \frac{p}{2}\log\left|\widehat{\mathbf{S}}_{\Gamma}\right| - \frac{1}{2}\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\alpha} - \sum_{j=1}^{K}\delta_j\beta_j\boldsymbol{\gamma}_j\right)^{\top}\widehat{\mathbf{S}}_{\Gamma}^{-1}\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\alpha} - \sum_{j=1}^{K}\delta_j\beta_j\boldsymbol{\gamma}_j\right) \\
&+ \sum_{j=1}^{K}\left[-\frac{p}{2}\log(2\pi) - \frac{p}{2}\log\left|\widehat{\mathbf{S}}_{\gamma_j}\right| - \frac{1}{2}\left(\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\right)^{\top}\widehat{\mathbf{S}}_{\gamma_j}^{-1}\left(\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\right)\right] \\
&- \frac{p}{2}\log(2\pi) - \frac{p}{2}\log\sigma_\alpha^2 - \frac{1}{2\sigma_\alpha^2}\sum_{i=1}^{p}\alpha_i^2 \\
&+ \sum_{j=1}^{K}\left[-\frac{p}{2}\log(2\pi) - \frac{p}{2}\log\sigma_\gamma^2 - \frac{1}{2\sigma_\gamma^2}\sum_{i=1}^{p}\gamma_{ij}^2\right] \\
&+ \sum_{j=1}^{K}\left[\delta_j\log\pi_j + (1-\delta_j)\log\left(1-\pi_j\right)\right] \\
=& \sum_{j=1}^{K}\left(\delta_j\beta_j\widehat{\boldsymbol{\Gamma}}^{\top}\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} - \delta_j\beta_j\boldsymbol{\alpha}^{\top}\widehat{\mathbf{S}}_{\Gamma}^{-1} - \delta_j\beta_j\sum_{t\neq j}^{K}\delta_t\beta_t\boldsymbol{\gamma}_t^{\top}\widehat{\mathbf{S}}_{\Gamma}^{-1} + \widehat{\boldsymbol{\gamma}}_j^{\top}\widehat{\mathbf{S}}_{\gamma_j}^{-1}\right)\boldsymbol{\gamma}_j - \sum_{j=1}^{K}\frac{1}{2}\boldsymbol{\gamma}_j^{\top}\left(\delta_j^2\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\gamma}_j \\
&+ \widehat{\boldsymbol{\Gamma}}^{\top}\widehat{\mathbf{S}}_{\Gamma}^{-1}\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^{\top}\left(\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \sigma_\alpha^{-2}\mathbf{I}\right)\boldsymbol{\alpha} \\
&- \frac{pK}{2}\log\sigma_\gamma^2 - \frac{p}{2}\log\sigma_{\boldsymbol{\alpha}}^2 + \sum_{j=1}^{K}\left[\delta_j\log\pi_j + (1-\delta_j)\log\left(1-\pi_j\right)\right] + \text{const.}
\end{aligned} \tag{4}
$$

We use mean field theory to approximate the posterior $q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) = \prod_{i=1}^{p} q\left(\alpha_i\right)\prod_{i=1}^{p}\prod_{j=1}^{K} q\left(\gamma_{ij} \mid \delta_j\right)\prod_{j=1} q(\delta_j)$. The posterior distribution of $\gamma_{ij} \sim \mathcal{N}\left(\mu_{ij}, \sigma_{ij}^2\right)$ when $\delta_j = 1$, where

$$
\begin{aligned}
-\frac{1}{2\sigma_{ij}^2} &= -\frac{\beta_j^2}{2}\frac{1}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_i}^2} - \frac{1}{2\widehat{\mathbf{s}}_{\gamma_i}^2} - \frac{1}{2\sigma_\gamma^2} \\
\frac{\mu_{ij}}{\sigma_{ij}^2} &= \beta_j\frac{\widehat{\Gamma}_i}{\widehat{\mathbf{s}}_{\Gamma_i}^2} - \frac{\beta_j}{\widehat{\mathbf{s}}_{\Gamma_i}}\cdot\frac{\langle\alpha_i\rangle}{\widehat{\mathbf{s}}_{\Gamma_i}} + \frac{\widehat{\gamma}_i}{\widehat{\mathbf{s}}_{\gamma_{ij}}^2} - \frac{\beta_j\sum_{t\neq j}^{K}\beta_t\left\langle\delta_t\gamma_{it}\right\rangle}{\widehat{\mathbf{s}}_{\Gamma_i}^2}
\end{aligned} \tag{5}
$$

when $\delta_j = 0$, we have $\gamma_{ij} \sim \mathcal{N}(\mu_{ij}^*, \sigma_{ij}^{2*})$, where

$$
\begin{aligned}
-\frac{1}{2\sigma_{ij}^{2*}} &= -\frac{1}{2\widehat{\mathbf{s}}_{\gamma_i}^2} - \frac{1}{2\sigma_\gamma^2} \\
\frac{\mu_{ij}^*}{\sigma_{ij}^{2*}} &= \frac{\widehat{\gamma}_{ij}}{\widehat{\mathbf{s}}_{\gamma_i}^2}
\end{aligned}
$$

The posterior distribution of $\alpha_i \sim \mathcal{N}\left(\widetilde{\mu}_i, \widetilde{\sigma}_i^2\right)$, where

$$\begin{aligned} -\frac{1}{2\widetilde{\sigma}_i^2} &= -\frac{1}{2\widehat{\mathrm{s}}_{\Gamma_i}^2} - \frac{1}{2\sigma_\alpha^2}, \\ \frac{\widetilde{\mu}_i}{\widetilde{\sigma}_i^2} &= \frac{\widehat{\Gamma}_i}{\widehat{\mathrm{s}}_{\Gamma_i}^2} - \frac{\sum_{j=1}^K \beta_j \left\langle \delta_j \gamma_{ij} \right\rangle}{\widehat{\mathrm{s}}_{\Gamma_i}^2}, \end{aligned} \tag{6}$$

where $\langle \gamma_k \rangle \stackrel{\text{def}}{=} E_q\left(\gamma_k\right)$. Similarly, $\boldsymbol{\alpha} \sim \mathcal{N}\left(\boldsymbol{\alpha} \mid \mu_\alpha, \boldsymbol{\Sigma}_\alpha\right)$, where $\mu_\alpha = \left(\widetilde{\mu}_1, \cdots, \widetilde{\mu}_p\right)^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_\alpha = \operatorname{diag}\left(\left\{\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_p^2\right\}\right)$.

We know that δ_j is a binary random variable, so we can let $\omega_j = q\left(\delta_j = 1\right)$, then the optimal variational distribution will be given by,

$$q\left(\gamma_{ij}, \alpha_i, \delta_j\right) = \left[\omega_j \mathcal{N}\left(\gamma_{ij} \mid \mu_{ij}, \sigma_{ij}^2\right)\right]^{\delta_j} \left[\left(1-\omega_j\right) \mathcal{N}\left(\gamma_{ij} \mid \mu_{ij}^*, \sigma_{ij}^{2*}\right)\right]^{1-\delta_j} \mathcal{N}\left(\alpha_i \mid \widetilde{\mu}_i, \widetilde{\sigma}_i^2\right)$$

In the M-step, we obtain the new updates for all the parameters $\boldsymbol{\theta}$ by setting the derivative of

$$E_q\left\{\log \operatorname{Pr}\left(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}_1, \cdots, \widehat{\boldsymbol{\gamma}}_K, \boldsymbol{\alpha} \mid \widehat{\mathbf{S}}_{\gamma_1}, \cdots, \widehat{\mathbf{S}}_{\gamma_K}, \widehat{\mathbf{S}}_\Gamma; \boldsymbol{\Theta}\right)\right\}$$

to zeros.

We have

$$\begin{aligned} &\mathbb{E}_q\left[\boldsymbol{\gamma}_j^\top\left(\delta_j^2\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\gamma}_j\right] \\ =&\mathbb{E}_q\left[\boldsymbol{\gamma}_j^\top\left(\delta_j^2\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\gamma}_j \mid \delta_j = 0\right]\mathbb{P}\left(\delta_j = 0\right) + \mathbb{E}_q\left[\boldsymbol{\gamma}_j^\top\left(\delta_j^2\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\gamma}_j \mid \delta_j = 1\right]\mathbb{P}\left(\delta_j = 1\right) \\ =&\left(1-\omega_j\right)\left\{\boldsymbol{\mu}_j^{*\mathrm{T}}\left(\widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\mu}_j^* + \operatorname{Tr}\left(\left(\widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\gamma_j}^*\right)\right\} \\ &+\omega_j\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\left(\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\mu}_j + \operatorname{Tr}\left(\left(\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\gamma_j}\right)\right\} \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}_q\left[\left(\delta_j\beta_j\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1} - \delta_j\beta_j\boldsymbol{\alpha}^\top\widehat{\mathbf{S}}_\Gamma^{-1} - \delta_j\beta_j\sum_{t\neq j}^K\delta_t\beta_t\boldsymbol{\gamma}_t^\top\widehat{\boldsymbol{S}}_\Gamma^{-1} + \widehat{\boldsymbol{\gamma}}_j^\top\widehat{\mathbf{S}}_{\gamma_j}^{-1}\right)\boldsymbol{\gamma}_j\right] \\ =&\left(\omega_j\beta_j\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1} - \omega_j\beta_j\boldsymbol{\mu}_\alpha^\top\widehat{\mathbf{S}}_\Gamma^{-1} - \omega_j\beta_j\sum_{t\neq j}^K\omega_t\beta_t\boldsymbol{\mu}_t^\top\widehat{\boldsymbol{S}}_\Gamma^{-1} + \widehat{\boldsymbol{\gamma}}_j^\top\widehat{\mathbf{S}}_{\gamma_j}^{-1}\right)\boldsymbol{\mu}_j \end{aligned}$$

and

$$\mathbb{E}_q\left[\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1}\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^\top\left(\widehat{\mathbf{S}}_\Gamma^{-1} + \sigma_\alpha^{-2}\mathbf{I}\right)\boldsymbol{\alpha}\right] = \widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\alpha}} - \left\{\boldsymbol{\mu}_\alpha^{\mathrm{T}}\left(\widehat{\mathbf{S}}_\Gamma^{-1} + \sigma_\alpha^{-2}\mathbf{I}\right)\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \operatorname{Tr}\left(\left(\widehat{\mathbf{S}}_\Gamma^{-1} + \sigma_\alpha^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\right)\right\}$$

and

$$\mathbb{E}\left[\delta_j \log \pi_j + \left(1-\delta_j\right)\log\left(1-\pi_j\right)\right] = \omega_j \log \pi_j + \left(1-\omega_j\right)\log\left(1-\pi_j\right)$$

We can evaluate the ELBO $\mathcal{L}(\Theta)$:

$$\mathcal{L}(\Theta) = E_q\left\{\log \operatorname{Pr}\left(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}_1, \cdots, \widehat{\boldsymbol{\gamma}}_K, \boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K, \boldsymbol{\delta} \mid \widehat{\mathbf{S}}_{\gamma_1}, \cdots, \widehat{\mathbf{S}}_{\gamma_K}, \widehat{\mathbf{S}}_\Gamma; \boldsymbol{\Theta}\right)\right\} - \mathbb{E}_q \log q\left(\boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K, \boldsymbol{\delta}\right)$$

Thus, we have

$$\begin{aligned} &E_q\left\{\log \operatorname{Pr}\left(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}_1, \cdots, \widehat{\boldsymbol{\gamma}}_K, \boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K, \boldsymbol{\delta} \mid \widehat{\mathbf{S}}_{\gamma_1}, \cdots, \widehat{\mathbf{S}}_{\gamma_K}, \widehat{\mathbf{S}}_\Gamma; \boldsymbol{\Theta}\right)\right\} \\ =&\sum_{j=1}^K\left(\omega_j\beta_j\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1} - \omega_j\beta_j\boldsymbol{\mu}_\alpha^\top\widehat{\mathbf{S}}_\Gamma^{-1} - \omega_j\beta_j\sum_{t\neq j}^K\omega_t\beta_t\boldsymbol{\mu}_t^\top\widehat{\boldsymbol{S}}_\Gamma^{-1} + \widehat{\boldsymbol{\gamma}}_j^\top\widehat{\mathbf{S}}_{\gamma_j}^{-1}\right)\boldsymbol{\mu}_j \\ &-\frac{1}{2}\sum_{j=1}^K\left(1-\omega_j\right)\left\{\boldsymbol{\mu}_j^{*\mathrm{T}}\left(\widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\mu}_j^* + \operatorname{Tr}\left(\left(\widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\gamma_j}^*\right)\right\} \\ &-\frac{1}{2}\sum_{j=1}^K\omega_j\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\left(\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\mu}_j + \operatorname{Tr}\left(\left(\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} + \widehat{\mathbf{S}}_{\gamma_j}^{-1} + \sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\gamma_j}\right)\right\} \\ &+\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\alpha}} - \left\{\boldsymbol{\mu}_\alpha^{\mathrm{T}}\left(\widehat{\mathbf{S}}_\Gamma^{-1} + \sigma_\alpha^{-2}\mathbf{I}\right)\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \operatorname{Tr}\left(\left(\widehat{\mathbf{S}}_\Gamma^{-1} + \sigma_\alpha^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\right)\right\} \\ &-\frac{pK}{2}\log\sigma_\gamma^2 - \frac{p}{2}\log\sigma_{\boldsymbol{\alpha}}^2 + \sum_{j=1}^K\left[\omega_j\log\pi_j + \left(1-\omega_j\right)\log\left(1-\pi_j\right)\right] + \text{ const.} \end{aligned}$$

Since

$$q\left(\gamma_{ij},\alpha_i,\delta_j\right)=\left[\omega_j\mathcal{N}\left(\gamma_{ij}\mid\mu_{ij},\sigma_{ij}^2\right)\right]^{\delta_j}\left[\left(1-\omega_j\right)\mathcal{N}\left(\gamma_{ij}\mid\mu_{ij}^*,\sigma_{ij}^{2*}\right)\right]^{1-\delta_j}\mathcal{N}\left(\alpha_i\mid\widetilde{\mu}_i,\widetilde{\sigma}_i^2\right)$$

The second term is

$$-\mathbb{E}_q\log q(\boldsymbol{\alpha},\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_K,\boldsymbol{\delta})=\sum_{j=1}^K\frac{1}{2}\omega_j\log\det\left(\boldsymbol{\Sigma}_{\gamma_j}\right)+\frac{1}{2}\left(1-\omega_j\right)\log\det\left(\boldsymbol{\Sigma}_{\gamma_j}^*\right)-\omega_j\log\omega_j-\left(1-\omega_j\right)\log\left(1-\omega_j\right)$$

Maximizing $\mathcal{L}$ w.r.t. ω_j is equivalent to set the following equation to zero

$$\begin{aligned}
&\left(\beta_j\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1}-\beta_j\boldsymbol{\mu}_\alpha^\top\widehat{\mathbf{S}}_\Gamma^{-1}-\beta_j\sum_{t\neq j}^K\omega_t\beta_t\boldsymbol{\mu}_t^\top\widehat{\boldsymbol{S}}_\Gamma^{-1}\right)\boldsymbol{\mu}_j\\
&+\frac{1}{2}\left\{\boldsymbol{\mu}_j^{*\mathrm{T}}\left(\widehat{\mathbf{S}}_{\gamma_j}^{-1}+\sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\mu}_j^*+\mathrm{Tr}\left(\left(\widehat{\mathbf{S}}_{\gamma_j}^{-1}+\sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\gamma_j}^*\right)\right\}\\
&-\frac{1}{2}\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\left(\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}+\widehat{\mathbf{S}}_{\gamma_j}^{-1}+\sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\mu}_j+\mathrm{Tr}\left(\left(\beta_j^2\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}+\widehat{\mathbf{S}}_{\gamma_j}^{-1}+\sigma_\gamma^{-2}\mathbf{I}\right)\boldsymbol{\Sigma}_{\gamma_j}\right)\right\}\\
&+\log\frac{\pi_j}{1-\pi_j}\\
&+\frac{1}{2}\log\det\left(\boldsymbol{\Sigma}_{\gamma_j}\right)-\frac{1}{2}\log\det\left(\boldsymbol{\Sigma}_{\gamma_j}^*\right)-\log\frac{\omega_j}{1-\omega_j}
\end{aligned}$$

which is

$$\boldsymbol{\mu}_j^\top\left(\boldsymbol{\Sigma}_{\gamma_j}^{-1}\boldsymbol{\mu}_j-\widehat{\mathbf{S}}_{\gamma_j}^{-1}\widehat{\boldsymbol{\gamma}}_j\right)+\frac{1}{2}\left\{\boldsymbol{\mu}_j^{*\mathrm{T}}\boldsymbol{\Sigma}_{\gamma_j}^{*-1}\boldsymbol{\mu}_j^*\right\}-\frac{1}{2}\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_{\gamma_j}^{-1}\boldsymbol{\mu}_j\right\}+\log\frac{\pi_j}{1-\pi_j}+\frac{1}{2}\log\det\left(\boldsymbol{\Sigma}_{\gamma_j}\right)-\frac{1}{2}\log\det\left(\boldsymbol{\Sigma}_{\gamma_j}^*\right)-\log\frac{\omega_j}{1-\omega_j}=0$$

So we obtain

$$\omega_j=\frac{1}{1+\exp\left(-\mathbf{b}_j\right)}$$

where

$$\mathbf{b}_j=\frac{1}{2}\boldsymbol{\mu}_j^\top\boldsymbol{\Sigma}_{\gamma_j}^{-1}\boldsymbol{\mu}_j+\frac{1}{2}\boldsymbol{\mu}_j^{*\ \mathrm{T}}\boldsymbol{\Sigma}_{\gamma_j}^{*-1}\boldsymbol{\mu}_j^*-\widehat{\boldsymbol{\gamma}}_j^\top\widehat{\mathbf{S}}_{\gamma_j}^{-1}\boldsymbol{\mu}_j+\log\frac{\pi_j}{1-\pi_j}+\frac{1}{2}\log\frac{\det\left(\boldsymbol{\Sigma}_{\gamma_j}\right)}{\det\left(\boldsymbol{\Sigma}_{\gamma_j}^*\right)}$$

M-step. We derive the updating equations for parameters $\beta_j,\pi_j,\sigma_\gamma^2,\sigma_\alpha^2$. We first derive the updating equation for β_j. The terms in $\partial\mathcal{L}\left(\Theta\right)/\partial\beta_j$ involving β_j are

$$\left(\omega_j\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1}-\omega_j\boldsymbol{\mu}_\alpha^\top\widehat{\mathbf{S}}_\Gamma^{-1}-\omega_j\sum_{t\neq j}^K\omega_t\beta_t\boldsymbol{\mu}_t^\top\widehat{\boldsymbol{S}}_\Gamma^{-1}\right)\boldsymbol{\mu}_j-\beta_j\omega_j\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\left(\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}\right)\boldsymbol{\mu}_j+\mathrm{Tr}\left(\left(\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}\right)\boldsymbol{\Sigma}_{\gamma_j}\right)\right\}=0$$

Finally, we have

$$\beta_j=\frac{\left(\widehat{\boldsymbol{\Gamma}}^\top\widehat{\mathbf{S}}_\Gamma^{-1}-\boldsymbol{\mu}_\alpha^\top\widehat{\mathbf{S}}_\Gamma^{-1}-\sum_{t\neq j}^K\omega_t\beta_t\boldsymbol{\mu}_t^\top\widehat{\boldsymbol{S}}_\Gamma^{-1}\right)\boldsymbol{\mu}_j}{\boldsymbol{\mu}_j^{\mathrm{T}}\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\mu}_j+\mathrm{Tr}\left(\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Sigma}_{\gamma_j}\right)}$$

By setting $\partial\mathcal{L}\left(\Theta\right)/\partial\sigma_\alpha^2=0$, we have

$$\frac{1}{2}\sigma_\alpha^{-4}\left\{\boldsymbol{\mu}_\alpha^{\mathrm{T}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}+\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\right)\right\}-\frac{p}{2}\sigma_\alpha^{-2}=0$$

So we have

$$\sigma_{\boldsymbol{\alpha}}^2=\frac{\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}+\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\right)}{p}$$

By setting $\partial\mathcal{L}\left(\Theta\right)/\partial\sigma_\gamma^2=0$, we have

$$\sum_{j=1}^K\left\{\frac{\sigma_\gamma^{-4}}{2}\left(1-\omega_j\right)\left\{\boldsymbol{\mu}_j^{*\mathrm{T}}\boldsymbol{\mu}_j^*+\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\gamma_j}^*\right)\right\}+\frac{\sigma_\gamma^{-4}}{2}\omega_j\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\mu}_j+\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\gamma_j}\right)\right\}\right\}-\frac{pK}{2}\sigma_\gamma^{-2}=0$$

So we have

$$\sigma_\gamma^2=\frac{\sum_{j=1}^K\left\{\left(1-\omega_j\right)\left\{\boldsymbol{\mu}_j^{*\mathrm{T}}\boldsymbol{\mu}_j^*+\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\gamma_j}^*\right)\right\}+\omega_j\left\{\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\mu}_j+\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\gamma_j}\right)\right\}\right\}}{pK}$$

By setting $\partial\mathcal{L}\left(\Theta\right)/\partial\sigma_\gamma^2=0$, we have

$$\pi_j=\omega_j$$