

Appendix A. Derivation of the proposed method

As in Q , the complete log-likelihood function Q has the form:

$$\begin{aligned}
Q(\Theta|\hat{\Theta}) &= E[L(\Pi, \Theta|Y)|Y, \hat{\Theta}] \\
&= \sum_{i=1}^N \sum_{k=1}^K \pi_{ik} \left\{ -\frac{1}{2} \log |\Lambda_k| - \frac{1}{2} (\mathbf{y}_i - \mathbf{x}_i - \boldsymbol{\mu}_k \mathbf{b}_i)^T \Lambda_k^{-1} (\mathbf{y}_i - \mathbf{x}_i - \boldsymbol{\mu}_k \mathbf{b}_i) - \frac{1}{2} \text{tr}(\Lambda_k^{-1} \hat{\Omega}_{ik}^{-1}) \right. \\
&\quad \left. - \frac{1}{2} \log |\Gamma_k| - \frac{1}{2} \text{tr} \{ \Gamma_k^{-1} (\hat{\Omega}_{i,k} + \hat{x}_{i,k} \hat{x}_{i,k}^T - \widehat{t_{i,k}} x_{i,k} \hat{\Delta}_k + \hat{t}_{i,k}^2 \hat{\Delta}_k \hat{\Delta}_k^T) \} \right\} \\
&\quad - \beta [K(s_i|\pi_i) + K(s_i|\pi_{\partial_i}) + H(s_i)] - \frac{1}{2} [K(q_i|z_i) + K(q_i|z_{\partial_i}) + H(q_i)]
\end{aligned} \tag{1}$$

Then we utilize EM algorithm to maximize the energy Q . In the E-step, we fix Θ and Π to maximize Q over s and q . In the M-step, we fix s and q to maximize Q over Θ and Π .

E-step: By fixing Θ and Π , we can optimize over s_i . The terms involving s_i in Q are:

$$\begin{aligned}
&K(s_i|\pi_i) + K(s_i|\pi_{\partial_i}) + H(s_i) \\
&= - \sum_{k=1}^K s_{ik} \log s_{ik} + \sum_{k=1}^K s_{ik} \log \pi_{ik} - \sum_{k=1}^K s_{ik} \log s_{ik} + \sum_{k=1}^K s_{ik} \log \pi_{\partial_{ik}} + \sum_{k=1}^K s_{ik} \log s_{ik} \\
&= - \sum_{k=1}^K s_{ik} \log s_{ik} + \sum_{k=1}^K s_{ik} \log (\pi_{ik} \pi_{\partial_{ik}})
\end{aligned} \tag{2}$$

The above formula is a negative KL-divergence which becomes zero when:

$$s_i \propto \pi_{ik} \pi_{\partial_{ik}} \tag{3}$$

By applying the similar derivation holds for q_i , we can get

$$q_i \propto z_{ik} z_{\partial_{ik}} \tag{4}$$

Therefore, we can get the updating functions for s_i and q_i as Eq.(2) and (4) in our

method.

M-step: By fixing s and q , we can maximize Q over Θ and Π . The terms involving Π and Θ are:

$$\begin{aligned} & \sum_{i=1}^N \left\{ \log \sum_{k=1}^K \{ \pi_{ik} p(\mathbf{x}_i | \Theta) \} \right\} \\ & - \beta [K(s_i | \pi_i) + \sum_{j \in \partial_i} K(s_j | \pi_{\partial_j})] - \frac{1}{2} [K(q_i | z_i) + \sum_{j \in \partial_i} K(q_j | z_{\partial_j})] \end{aligned} \quad (5)$$

First of all, let us consider the derivation over z_i , then the terms involving only z_i are:

$$\begin{aligned} & - \frac{1}{2} [K(q_i | z_i) + \sum_{j \in \partial_i} K(q_j | z_{\partial_j})] \\ & = - \frac{1}{2} \left[\sum_{k=1}^K q_{ik} \log q_{ik} - \sum_{k=1}^K q_{ik} \log z_{ik} + \sum_{j \in \partial_i, j \neq i} \left(\sum_{k=1}^K q_{jk} \log q_{jk} - \sum_{k=1}^K q_{jk} \log z_{\partial_j k} \right) \right] \end{aligned} \quad (6)$$

By ignoring the terms independent of z_{ik} , we get

$$- \frac{1}{2} \left[- \sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{j \in \partial_i, j \neq i} \sum_{k=1}^K q_{jk} \log z_{\partial_j k} \right] \quad (7)$$

Where

$$z_{\partial_j} = \sum_{m \in \partial_j, m \neq j} \alpha_{jm} z_i = \alpha_{jm} z_i + \sum_{m \in \partial_j, m \neq i, j} \alpha_{jm} z_m \quad (8)$$

To make the M-step tractable, we using Jensens inequality to bound terms in Eq.(8):

$$\log z_{\partial_j k} = \log \sum_{m \in \partial_j, m \neq j} \alpha_{jm} z_{mk} \geq \alpha_{jm} \log z_{ik} + \log \sum_{m \in \partial_j, m \neq i, j} \alpha_{jm} z_m \quad (9)$$

Since $\alpha_{ji} = \alpha_{ij}$, By combining Eq. (7), (8) and (9), we obtain:

$$\begin{aligned}
& \frac{1}{2} \left[- \sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{j \in \partial_i, j \neq i} \sum_{k=1}^K q_{jk} \log z_{\partial_j k} \right] \\
&= \frac{1}{2} \left[- \sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{j \in \partial_i, j \neq i} \sum_{k=1}^K q_{jk} \log \left(\sum_{m \in \partial_j, m \neq j} \alpha_{jm} z_{mk} \right) \right] \\
&\geq \frac{1}{2} \left[\sum_{k=1}^K q_{ik} \log z_{ik} + \sum_{j \in \partial_i, j \neq i} \left(\sum_{k=1}^K q_{jk} (\alpha_{ji} \log z_{ik}) + \sum_{m \in \partial_j, m \neq j} \alpha_{jm} \log z_{mk} \right) \right] \\
&= \frac{1}{2} \left[\sum_{k=1}^K q_{ik} \log z_{ik} + \sum_{k=1}^K \sum_{j \in \partial_i, j \neq i} q_{jk} (\alpha_{ji} \log z_{ik}) + \sum_{j \in \partial_i, j \neq i} \sum_{m \in \partial_j, m \neq j} \alpha_{jm} \log z_{mk} \right]
\end{aligned} \tag{10}$$

By only preserving the terms involving q_i , then the remaining terms of formula (16) are:

$$\begin{aligned}
& \frac{1}{2} \left[- \sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{k=1}^K \sum_{j \in \partial_i, j \neq i} q_{jk} \log q_{jk} (\alpha_{ji} \log z_{ik}) \right] \\
&= \frac{1}{2} \left[- \sum_{k=1}^K q_{ik} \log z_{ik} - \sum_{k=1}^K \sum_{j \in \partial_i, j \neq i} \alpha_{ji} q_{jk} \log z_{ik} \right] \\
&= \frac{1}{2} \left[\sum_{k=1}^K (q_{ik} + \sum_{j \in \partial_i, j \neq i} \alpha_{ji} q_{jk}) \log z_{ik} \right] \\
&\implies \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik}
\end{aligned} \tag{11}$$

Where the distribution q_{∂_i} is

$$q_{\partial_i} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} q_j \tag{12}$$

By applying the similar derivation holds for π_i , we can get:

$$\beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \tag{13}$$

Where the distribution s_{∂_i} is:

$$s_{\partial_i} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} s_j \quad (14)$$

Consequently, the lower bound of complete log-likelihood function Q involving the posterior z_i and prior π_i becomes:

$$\log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik} \quad (15)$$

In Eq.(15), $\frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) = \frac{1}{2} \sum_{k=1}^K q_{ik} + \frac{1}{2} \sum_{k=1}^K q_{\partial_i k} = 1$. By expanding the posterior z_{ik} , we find that maximizing

$$\begin{aligned} & \log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik} \\ &= \log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log z_{ik} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \\ &= \log \sum_{k=1}^K \{\pi_{ik} p(\mathbf{x}_i | \Theta)\} + \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log \left\{ \frac{\pi_{ik} p(\mathbf{x}_i | \Theta)}{\sum_{k=1}^K \pi_{ik} p(\mathbf{x}_i | \Theta)} \right\} + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \\ &= \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log (\pi_{ik} p(\mathbf{x}_i | \Theta)) + \beta \sum_{k=1}^K (s_{ik} + s_{\partial_i k}) \log \pi_{ik} \end{aligned} \quad (16)$$

is equivalent to maximizing:

$$\frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_i k}) \log p(\mathbf{x}_i | \Theta) + \sum_{k=1}^K \left(\frac{1}{2} (q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_i k}) \right) \log \pi_{ik} \quad (17)$$

Then, the Lagranges multiplier is used to enforce the constraint $\sum_{k=1}^K \pi_{ik} = 1$ for

each data point, and we can easily get the updating function for the prior π_{ik} :

$$\hat{\pi}_{ik} = \frac{1}{1+2\beta} \left(\frac{1}{2}(q_{ik} + q_{\partial_{ik}}) + \beta(s_{ik} + s_{\partial_{ik}}) \right) \quad (18)$$

Similarly, we obtain the following update equations for $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$, and . The energy function can be rewritten as:

$$\begin{aligned} Q^* &= \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_{ik}}) \log p(\mathbf{x}_i | \boldsymbol{\Theta}) + \sum_{k=1}^K \left(\frac{1}{2}(q_{ik} + q_{\partial_{ik}}) + \beta(s_{ik} + s_{\partial_{ik}}) \right) \log \pi_{ik} \\ &= \frac{1}{2} \sum_{k=1}^K (q_{ik} + q_{\partial_{ik}}) \left\{ -\frac{1}{2} \log |\widetilde{\Sigma}_k| - \frac{1}{2} (\mathbf{y}_i - \mathbf{x}_i - \boldsymbol{\mu}_k \mathbf{b}_i)^T \widetilde{\Sigma}_k^{-1} (\mathbf{y}_i - \mathbf{x}_i - \boldsymbol{\mu}_k \mathbf{b}_i) \right. \\ &\quad \left. - \frac{1}{2} \log |\Gamma_k| - \frac{1}{2} \text{tr} \{ \Gamma_k^{-1} (\widehat{\Omega}_{i,k} + \widehat{x}_{i,k} \widehat{x}_{i,k}^T - \widehat{t_{i,k}} x_{i,k} \widehat{\Delta}_k + \widehat{t_{i,k}}^2 \widehat{\Delta}_k \widehat{\Delta}_k^T) \} \right. \\ &\quad \left. + \sum_{k=1}^K \left(\frac{1}{2}(q_{ik} + q_{\partial_{ik}}) + \beta(s_{ik} + s_{\partial_{ik}}) \right) \log \pi_{ik} \right. \end{aligned} \quad (19)$$

Let $\frac{\partial Q^*}{\partial \boldsymbol{\mu}_k} = 0$, $\frac{\partial Q^*}{\partial \boldsymbol{\Delta}_k} = 0$, $\frac{\partial Q^*}{\partial \widetilde{\Sigma}_k^{-1}} = 0$ and $\frac{\partial Q^*}{\partial \Gamma_k^{-1}} = 0$, we can obtain:

$$\hat{\pi}_{ik} = \frac{1}{1+2\beta} \left(\frac{1}{2}(q_{ik} + q_{\partial_{ik}}) + \beta(s_{ik} + s_{\partial_{ik}}) \right) \quad (20)$$

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}}) (y_i - \widehat{x}_{i,k})}{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}})} \quad (21)$$

$$\widehat{\Delta}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}}) \widehat{t_{i,k}} x_{i,k}}{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}}) \widehat{t_{i,k}}^2} \quad (22)$$

$$\widehat{\Lambda}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}}) ((y_i - \mu_k b_i - \widehat{x}_{i,k})(y_i - \mu_k b_i - \widehat{x}_{i,k})^T + \widehat{\Omega}_{i,k})}{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}})} \quad (23)$$

$$\widehat{\Gamma}_k = \frac{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}}) (\widehat{\Omega}_{i,k} + \widehat{x}_{i,k} \widehat{x}_{i,k}^T - 2 \widehat{t_{i,k}} x_{i,k} \Delta_k + \widehat{t_{i,k}}^2 \Delta_k \Delta_k^T)}{\sum_{i=1}^N (q_{ik} + q_{\partial_{ik}})} \quad (24)$$

$$\widehat{b}_i = A^T G(i), A = E^{-1} V \quad (25)$$

$$\text{where } E = \sum_{i=1}^N G(i)G(i)^T J_1, \quad V = \sum_{i=1}^N G(i)J_2, \quad J_1 = \sum_{i=1}^K \hat{Z}_{ik} \boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\mu}_k, \quad J_2 = \sum_{i=1}^K \hat{Z}_{ik} y_i \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\mu}_k - \hat{Z}_{ik} \hat{x}_{ik} \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\mu}_k, \quad \boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k + \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^T. \quad \lambda_k = \frac{\boldsymbol{\Sigma}_k^{-1/2} \boldsymbol{\Delta}_k}{\sqrt{1 - \boldsymbol{\Delta}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_k}}.$$