1 Appendix A

The Q function of the Eq.(28) can be define as:

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

Then we utilize EM algorithm to maximize the energy Q, In the E-step, we fix $\mathbf{\Theta}$ and $\mathbf{\Pi}$ to maximize Q over s and q. In the M-step, we fix s and q to maximize Q over $\mathbf{\Theta}$ and $\mathbf{\Pi}$.

E-step: By fixing $\mathbf{\Theta}$ and $\mathbf{\Pi}$, we can optimize over s_i. The terms involving s_i in Q are:

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

The above formula is a negative KL-divergence which becomes zero when[1]:

$$s_i \propto \pi_{ik}\pi_{\partial_i k} \tag{3}$$

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

$$q_i \propto z_{ik}z_{\partial_i k} \tag{4}$$

Therefore, we can get the updating functions for s_i and q_i as Eq.(31) and (32) in our manuscript.

M-step: By fixing s and q, we can maximize Q over $\mathbf{\Theta}$ and $\mathbf{\Pi}$. The terms involving $\mathbf{\Pi}$ and $\mathbf{\Theta}$ are:

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

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

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

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

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

Where

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

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

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

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

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

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

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

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:

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

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

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

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

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

is equivalent to maximizing:

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

Then, the Lagrange's 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}:

$$\widehat{\pi}_{ik} = \frac{1}{1 + 2\beta} (\frac{1}{2}(q_{ik} + q_{\partial_i k}) + \beta (s_{ik} + s_{\partial_{ik}})) \tag{18}$$

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

$$
\begin{aligned}
Q^* &= \frac{1}{2}\sum_{K=1}^{K}(q_{ik}+q_{\partial_i k})\log p(\mathbf{x}_i|\mathbf{\Theta}) + \sum_{K=1}^{K}(\frac{1}{2}(q_{ik}+q_{\partial_i k})+\beta(s_{ik}+s_{\partial_i k}))\log \pi_{ik} \\
&= \frac{1}{2}\sum_{K=1}^{K}(q_{ik}+q_{\partial_i k})\{-\frac{1}{2}\log|\mathbf{\Lambda}_k| - \frac{1}{2}(\mathbf{y}_i-\mathbf{x}_i-\boldsymbol{\mu}_k)^T\mathbf{\Lambda}_k^{-1}(\mathbf{y}_i-\mathbf{x}_i-\boldsymbol{\mu}_k) \\
&- \frac{1}{2}\log|\Gamma_k| - \frac{1}{2}tr\{\Gamma_k^{-1}(\widehat{\Omega}_{i,k}+\widehat{x}_{i,k}\widehat{x}_{i,k}^T - \widehat{t_{i,k}x}_{i,k}\widehat{\mathbf{\Delta}}_k + \widehat{t}_{i,k}^2\widehat{\mathbf{\Delta}}_k\widehat{\mathbf{\Delta}}_k^T)\} \\
&+ \sum_{k=1}^{K}(\frac{1}{2}(q_{ik}+q_{\partial_i k})+\beta(s_{ik}+s_{\partial_i k})))\log \pi_{ik}
\end{aligned} \tag{19}
$$

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

$$\widehat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})(y_i-\widehat{x}_{i,k})}{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})} \tag{20}$$

$$\widehat{\mathbf{\Delta}}_k = \frac{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})\widehat{t_{i,k}x}_{i,k}}{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})\widehat{t}_{i,k}^2} \tag{21}$$

$$\widehat{\mathbf{\Lambda}}_k = \frac{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})((y_i-\mu_k-\widehat{x}_{i,k})(y_i-\mu_k-\widehat{x}_{i,k})^T+\widehat{\Omega}_{i,k}}{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})} \tag{22}$$

$$\widehat{\mathbf{\Gamma}}_k = \frac{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})(\widehat{\Omega}_{i,k}+\widehat{x}_{i,k}\widehat{x}_{i,k}^T - 2\widehat{t_{i,k}x}_{i,k}\Delta_k + \widehat{t}_{i,k}^2\Delta_k\Delta_k^T)}{\sum_{i=1}^N (q_{ik}+q_{\partial_i k})} \tag{23}$$

Where $\mathbf{\Sigma}_k = \mathbf{\Gamma}_k + \mathbf{\Delta}_k\mathbf{\Delta}_k^T$. $\boldsymbol{\lambda}_k = \frac{\mathbf{\Sigma}_k^{-1/2}\mathbf{\Delta}_k}{\sqrt{1-\mathbf{\Delta}_k^T\mathbf{\Sigma}_k^{-1}\mathbf{\Delta}_k}}$.

References

[1] Zexuan Ji, Yubo Huang, Quansen Sun, and Guo Cao. A spatially constrained generative asymmetric gaussian mixture model for image segmentation. *Journal of Visual Communication and Image Representation*, 40:611 – 626, 2016.