Supplementary Material of "A Gaussian Mixture Model for Discovering Latent Group Structures in Classification Problems with Multiple Classes"

Hong Chang[1,a], Xuetong Li[2,c,*], Ke Xu[3,d], Hansheng Wang[1,b]

[1] *Guanghua School of Management, Peking University,*

[a] *changhong@stu.pku.edu.cn;* [b] *hansheng@gsm.pku.edu.cn*

[2] *School of Mathematics and Statistics, Xi'an Jiaotong University,*

[c] *xtong_li@xjtu.edu.cn*

[3] *School of Statistics, University of International Business and Economics,*

[d] *xk@uibe.edu.cn*

APPENDIX A

Appendix A.1: Proof of the Log-Likelihood Function

In this section, we prove the equation (2) in Section 2.2. It is worth noting that $\mathbf{X}_{(\mathcal{I}_k)}$s with different class labels k are independent. Then, we have $f_{\boldsymbol{\theta}}(\mathbf{X}|\boldsymbol{Y},\mathbb{B}) = P(\mathbf{X}|\boldsymbol{Y},\mathbb{B}) = \prod_{k=1}^{K} P(\mathbf{X}_{(\mathcal{I}_k)}|\boldsymbol{Y}_{(\mathcal{I}_k)},\mathbb{B})$. Recall that $\phi_{\boldsymbol{\nu}_m,\boldsymbol{\Omega}}(\cdot)$ is the probability density function of the random mean vector $\boldsymbol{\mu}_{km}$ given $Y_i = k$ and $Z_i = m$. Additionally, $\phi_{\boldsymbol{\mu}_{km},\boldsymbol{\Sigma}}(\cdot)$ is the probability density function of $\boldsymbol{X}_i$, given $Y_i = k$, $Z_i = m$ and $\boldsymbol{\mu}_{km}$. Since $\boldsymbol{X}_i$s with different is are independently generated, conditional on $Y_i = k$, $Z_k = m$, and $\mathbb{B}$, the conditional likelihood function can be proved as

$$
\begin{aligned}
f_{\boldsymbol{\theta}}\big(\mathbf{X}\big|\boldsymbol{Y},\mathbb{B}\big) &= \prod_{k=1}^{K}\left\{\sum_{m=1}^{M}\alpha_m P\Big(\mathbf{X}_{(\mathcal{I}_k)}\Big|\boldsymbol{Y}_{(\mathcal{I}_k)}, Z_k = m, \mathbb{B}\Big)\right\} \\
&= \prod_{k=1}^{K}\left\{\sum_{m=1}^{M}\alpha_m \prod_{i\in\mathcal{I}_k}\phi_{\boldsymbol{\mu}_{km},\boldsymbol{\Sigma}}(\boldsymbol{X}_i)\right\}. \qquad \text{(A.1)}
\end{aligned}
$$

*Xuetong Li is the corresponding author.

Considering the randomness of the random variable vector set $\mathbb{B}$ for (A.1), we then compute the marginal likelihood function (A.1) as

$$f_{\boldsymbol{\theta}}(\mathbf{X}|\boldsymbol{Y}) = \prod_{k=1}^{K}\left[\sum_{m=1}^{M}\alpha_m\int\Big\{\prod_{i\in\mathcal{I}_k}\phi_{\boldsymbol{\mu}_{km},\boldsymbol{\Sigma}}(\boldsymbol{X}_i)\phi_{\boldsymbol{\nu}_m,\boldsymbol{\Omega}}(\boldsymbol{\mu}_{km})\Big\}d\boldsymbol{\mu}_{km}\right]$$
$$= \prod_{k=1}^{K}\left[\sum_{m=1}^{M}\alpha_m\int(2\pi)^{-p(N_k+1)/2}\big|\boldsymbol{\Sigma}\big|^{-N_k/2}\big|\boldsymbol{\Omega}\big|^{-1/2}\exp\Big\{-2^{-1}g\big(\boldsymbol{\mu}_{km}\big)\Big\}d\boldsymbol{\mu}_{km}\right], \quad \text{(A.2)}$$

where $g(\boldsymbol{\mu}_{km}) = \sum_{i\in\mathcal{I}_k}(\boldsymbol{X}_i-\boldsymbol{\mu}_{km})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i-\boldsymbol{\mu}_{km}) + (\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m)^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m) = \sum_{i\in\mathcal{I}_k}(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k)+g_1(\boldsymbol{\mu}_{km})$ and $g_1(\boldsymbol{\mu}_{km}) = (\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m)^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m)+N_k(\overline{\boldsymbol{X}}_k-\boldsymbol{\mu}_{km})^\top\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}_k-\boldsymbol{\mu}_{km})$. By a simple calculation, we can prove that $g_1(\boldsymbol{\mu}_{km}) = (\boldsymbol{\mu}_{km}-\boldsymbol{\mu}_k^*)^\top\boldsymbol{\Sigma}_k^{*-1}(\boldsymbol{\mu}_{km}-\boldsymbol{\mu}_k^*)-\boldsymbol{\mu}^{*\top}\boldsymbol{\Sigma}_k^{*-1}\boldsymbol{\mu}_k^*+N_k\overline{\boldsymbol{X}}_k\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k+\boldsymbol{\nu}_m^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}_m$, where $\boldsymbol{\Sigma}_k^{*-1} = N_k\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1}$ and $\boldsymbol{\mu}_k^* = \boldsymbol{\Sigma}^*(N_k\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{X}}_k+\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}_m)$. Therefore, $g(\boldsymbol{\mu}_{km}) = (\boldsymbol{\mu}_{km}-\boldsymbol{\mu}_k^*)^\top\boldsymbol{\Sigma}_k^{*-1}(\boldsymbol{\mu}_{km}-\boldsymbol{\mu}_k^*)+h(\mathbf{X}_{(\mathcal{I}_k)})$, where $h(\mathbf{X}_{(\mathcal{I}_k)}) = \sum_{i\in\mathcal{I}_k}\boldsymbol{X}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i+\boldsymbol{\nu}_m^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}_m-\boldsymbol{\mu}_k^{*\top}\boldsymbol{\Sigma}_k^{*-1}\boldsymbol{\mu}_k^*$. Combining the result with (A.2), the following equation can be given as

$$f_{\boldsymbol{\theta}}(\mathbf{X}|\boldsymbol{Y}) = \prod_{k=1}^{K}\left[\sum_{m=1}^{M}\alpha_m(2\pi)^{-pN_k/2}\big|\boldsymbol{\Sigma}\big|^{-N_k/2}\big|\boldsymbol{\Omega}\big|^{-1/2}\big|\boldsymbol{\Sigma}_k^*\big|^{1/2}\exp\Big\{-2^{-1}h\big(\mathbf{X}_{(\mathcal{I}_k)}\big)\Big\}\right.$$
$$\left.\int(2\pi)^{-p/2}\big|\boldsymbol{\Sigma}_k^*\big|^{1/2}\exp\Big\{-2^{-1}\big(\boldsymbol{\mu}_{km}-\boldsymbol{\mu}_k^*\big)^\top\boldsymbol{\Sigma}_k^{*-1}\big(\boldsymbol{\mu}_{km}-\boldsymbol{\mu}_k^*\big)\Big\}d\boldsymbol{\mu}_{km}\right]. \quad \text{(A.3)}$$

Since the integral part in (A.3) results in 1, we then have $f_{\boldsymbol{\theta}}(\mathbf{X}|\boldsymbol{Y}) = \prod_{k=1}^{K}[\sum_{m=1}^{M}\alpha_m (2\pi)^{-pN_k/2}|\boldsymbol{\Sigma}|^{-N_k/2}|\boldsymbol{\Omega}|^{-1/2}|\boldsymbol{\Sigma}_k^*|^{1/2}\exp\{-2^{-1}h(\mathbf{X}_{(\mathcal{I}_k)})\}]$. Therefore, the simplified log-likelihood function can be obtained by taking the logarithm of $f_{\boldsymbol{\theta}}(\mathbf{X}|\boldsymbol{Y})$ and discarding the irrelevant constant, as shown in the following.

$$\mathcal{L}(\boldsymbol{\theta}) = \log\Big\{f_{\boldsymbol{\theta}}\big(\mathbf{X}|\boldsymbol{Y}\big)\Big\} = -2^{-1}N\log\big|\boldsymbol{\Sigma}\big| - 2^{-1}K\log\big|\boldsymbol{\Omega}\big| - \sum_{k=1}^{K}2^{-1}\log\big|\boldsymbol{\Sigma}_k^{*-1}\big| +$$
$$\sum_{k=1}^{K}\log\left[\sum_{m=1}^{M}\alpha_m\exp\Big\{-2^{-1}\Big(\sum_{i\in\mathcal{I}_k}\boldsymbol{X}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i+\boldsymbol{\nu}_m^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}_m-\boldsymbol{\mu}_k^{*\top}\boldsymbol{\Sigma}_k^{*-1}\boldsymbol{\mu}_k^*\Big)\Big\}\right].$$

Next, with simple rearrangement, we can obtain the log-likelihood function given in equation (2).

Appendix A.2: The Derivation of the Mapping Functions

The conclusion about the first order derivative can be proved in a total of three steps. In the first step, we should start with a careful calculation of the first order derivative. Next, in the second step, a mapping function can be obtained with the gradient condition $\dot{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}) = \mathbf{0}$. In the last step, we can compute the expectation and the covariance of $\dot{\mathcal{L}}(\widehat{\boldsymbol{\theta}})$.

Step 1. Following Boldea and Magnus (2009), the first order derivative can be computed by using differentiation. Recall that

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) =& -2^{-1}N\log\big|\boldsymbol{\Sigma}\big| - 2^{-1}K\log\big|\boldsymbol{\Omega}\big| - \sum_{k=1}^{K} 2^{-1}\log\big|\boldsymbol{\Sigma}_k^{*-1}\big| \\ &+ \sum_{k=1}^{K}\log\left(\sum_{m=1}^{M}\alpha_m\exp\left[-2^{-1}\Big\{\sum_{i\in\mathcal{I}_k}\big(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k\big)^\top\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k\big)\right.\right. \\ &\left.\left. + \big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)^\top\big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)\Big\}\right]\right),\end{aligned}$$

where $\boldsymbol{\Sigma}_k^{*-1} = N_k\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^{-1}$. Therefore, the derivative of $\mathcal{L}(\boldsymbol{\theta})$ can be given by

$$d\mathcal{L}(\boldsymbol{\theta}) = d\mathcal{L}_1(\boldsymbol{\theta}) + d\mathcal{L}_2(\boldsymbol{\theta}) + d\mathcal{L}_3(\boldsymbol{\theta}) + d\mathcal{L}_4(\boldsymbol{\theta}), \tag{A.4}$$

where $d\mathcal{L}_1(\boldsymbol{\theta}) = -2^{-1}Nd\log|\boldsymbol{\Sigma}|$, $d\mathcal{L}_2(\boldsymbol{\theta}) = -2^{-1}Kd\log|\boldsymbol{\Omega}|$, $d\mathcal{L}_3(\boldsymbol{\theta}) = -\sum_{k=1}^{K}2^{-1}d\log|\boldsymbol{\Sigma}_k^{*-1}|$, $d\mathcal{L}_4(\boldsymbol{\theta}) = \sum_{k=1}^{K}d\log(\sum_{m=1}^{M}\tau_{mk})$, and $\tau_{mk} = \alpha_m\exp[-2^{-1}\{\sum_{i\in\mathcal{I}_k}(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k) + (\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)^\top(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)\}]$.

By a straightforward matrix derivative calculation (Boldea and Magnus, 2009), we can obtain that $d\mathcal{L}_1(\boldsymbol{\theta}) = -2^{-1}N\mathrm{tr}(\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}) = -2^{-1}N\{\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\}^\top\mathbf{D}d\mathrm{vech}(\boldsymbol{\Sigma})$ and $d\mathcal{L}_2(\boldsymbol{\theta}) = -2^{-1}K\mathrm{tr}(\boldsymbol{\Omega}^{-1}d\boldsymbol{\Omega}) = -2^{-1}K\{\mathrm{vec}(\boldsymbol{\Omega}^{-1})\}^\top\mathbf{D}d\mathrm{vech}(\boldsymbol{\Omega})$. We next consider $d\mathcal{L}_3(\boldsymbol{\theta}) = \sum_{k=1}^{K}2^{-1}d\log|\boldsymbol{\Sigma}_k^{*}|$, where $\boldsymbol{\Sigma}_k^{*} = \boldsymbol{\Omega}(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}(N_k^{-1}\boldsymbol{\Sigma})$. We have $d\log|\boldsymbol{\Sigma}_k^{*}| = d\log|\boldsymbol{\Omega}| + d\log|N_k^{-1}\boldsymbol{\Sigma}| - d\log|N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}| = \mathrm{tr}[\{\boldsymbol{\Omega}^{-1} - (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}d\boldsymbol{\Omega}] +$

$\mathrm{tr}[\{\boldsymbol{\Sigma}^{-1} - N_k^{-1}(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\}d\boldsymbol{\Sigma}]$. Therefore, it can be verified that

$$\begin{aligned} d\mathcal{L}_3(\boldsymbol{\theta}) =& 2^{-1}\sum_{k=1}^K \mathrm{vec}\Big\{\boldsymbol{\Omega}^{-1} - \big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Omega}) \\ &+ 2^{-1}\sum_{k=1}^K \mathrm{vec}\Big\{\boldsymbol{\Sigma}^{-1} - N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Sigma}). \end{aligned}$$

We next study $d\mathcal{L}_4(\boldsymbol{\theta}) = \sum_{k=1}^K d\log(\sum_{m=1}^M \tau_{mk})$, where

$$d\log\left(\sum_{m=1}^M \tau_{mk}\right) = \sum_{m=1}^M \frac{d\tau_{mk}}{\sum_{m'=1}^M \tau_{m'k}} = \sum_{m=1}^M \frac{\tau_{mk}}{\sum_{m'=1}^M \tau_{m'k}} d\log\tau_{mk} = \sum_{m=1}^M \pi_{(\mathcal{I}_k)}^m d\log\tau_{mk}.$$

Note that $\pi_{(\mathcal{I}_k)}^m = (\sum_{m'=1}^M \tau_{m'k})^{-1}\tau_{mk}$ can be written as

$$\begin{aligned} \pi_{(\mathcal{I}_k)}^m =& \Big[\alpha_m \exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)^\top \big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)\Big\}\Big] \\ &\times \Big[\sum_{m'=1}^M \alpha_{m'} \exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_{m'}\big)^\top \big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_{m'}\big)\Big\}\Big]^{-1}. \end{aligned}$$

Next, we study $d\mathcal{L}_4(\boldsymbol{\theta}) = \sum_{k=1}^K \sum_{m=1}^M \pi_{(\mathcal{I}_k)}^m d\log\tau_{mk}$. To this end, we compute $d\log\tau_{mk} = d\log\alpha_m - 2^{-1}\sum_{i\in\mathcal{I}_k}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)^\top d(\boldsymbol{\Sigma}^{-1})(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k) - 2^{-1}d(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)^\top (N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)$. Next, the first term is given by $d\log\alpha_m = \alpha_m^{-1}d\alpha_m$ with $1 \le m \le M-1$, and $d\log\alpha_M = -\sum_{m=1}^{M-1}(1 - \sum_{m'=1}^{M-1}\alpha_{m'})^{-1}d\alpha_m$. The second term can be given as

$$\begin{aligned} &-2^{-1}\sum_{i\in\mathcal{I}_k}\big(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k\big)^\top d\big(\boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k\big) \\ =& 2^{-1}\sum_{i\in\mathcal{I}_k}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)^\top \boldsymbol{\Sigma}^{-1} d\big(\boldsymbol{\Sigma}\big)\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k) \\ =& 2^{-1}\mathrm{tr}\Big(\sum_{i\in\mathcal{I}_k}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top d\boldsymbol{\Sigma}\Big) = 2^{-1}\Big\{\mathrm{vec}\Big(\sum_{i\in\mathcal{I}_k}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\Big)\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Sigma}), \end{aligned}$$

where $\boldsymbol{c}_{ik} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)$. We then compute the last term of $d\log\tau_{mk}$ as

$$\begin{aligned} &-2^{-1}d\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)^\top \big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big) \\ =& 2^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)^\top \big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1} d\big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)\big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big) \end{aligned}$$

$$
\begin{aligned}
&+ \big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)^\top \big(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1} d\boldsymbol{\nu}_m \\
=&\boldsymbol{b}_{km}^\top d\boldsymbol{\nu}_m + \big(2N_k\big)^{-1}\mathrm{tr}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top d\boldsymbol{\Sigma}\Big) + 2^{-1}\mathrm{tr}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top d\boldsymbol{\Omega}\Big) \\
=&\boldsymbol{b}_{km}^\top d\boldsymbol{\nu}_m + \big(2N_k\big)^{-1}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Sigma}) + 2^{-1}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Omega}),
\end{aligned}
$$

where $\boldsymbol{b}_{km} = (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)$. By combing the analytical results of the above three terms, we can obtain that

$$
\begin{aligned}
d\log\tau_{mk} =&\alpha_m^{-1} d\alpha_m + \boldsymbol{b}_{km}^\top d\boldsymbol{\nu}_m + 2^{-1}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Omega}) \\
&+ 2^{-1}\Big\{\mathrm{vec}\Big(\sum_{i\in\mathcal{I}_k} \boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\Big) + N_k^{-1}\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}^\top \mathbf{D} d\mathrm{vech}(\boldsymbol{\Sigma}),
\end{aligned}
$$

where $1 \le m \le M-1$. Similarly, we can obtain $d\log\tau_{Mk}$ by replacing $\alpha_m^{-1}d\alpha_m$ with $-\sum_{m=1}^{M-1}(1-\sum_{m'=1}^{M-1}\alpha_{m'})^{-1}d\alpha_m$.

Combing the above results, we can evaluate $d\mathcal{L}(\boldsymbol{\theta}) = d\mathcal{L}_1(\boldsymbol{\theta}) + d\mathcal{L}_2(\boldsymbol{\theta}) + d\mathcal{L}_3(\boldsymbol{\theta}) + \sum_{k=1}^K\sum_{m=1}^M \pi_{(\mathcal{I}_k)}^m d\log\tau_{mk}$ as follows. We then have $d\mathcal{L}(\boldsymbol{\theta}) = \dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})^\top d\boldsymbol{\alpha} + \dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta})^\top d\nu + \dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta})^\top d\boldsymbol{\Omega} + \dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})^\top d\boldsymbol{\Sigma}$, where

$$
\begin{aligned}
\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) =&\sum_{k=1}^K\sum_{m=1}^M \pi_{(\mathcal{I}_k)}^m \widetilde{\boldsymbol{\alpha}}_m, \quad \dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta}) = (\sum_{k=1}^K \pi_{(\mathcal{I}_k)}^1 \boldsymbol{b}_{k1}^\top, \ldots, \sum_{k=1}^K \pi_{(\mathcal{I}_k)}^M \boldsymbol{b}_{kM}^\top)^\top, \\
\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) =&2^{-1}\sum_{k=1}^K\sum_{m=1}^M \pi_{(\mathcal{I}_k)}^m \mathbf{D}^\top\Big[\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top) - \mathrm{vec}\big\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\big\}\Big], \\
\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) =&2^{-1}\sum_{k=1}^K\sum_{m=1}^M \pi_{(\mathcal{I}_k)}^m \mathbf{D}^\top\Bigg\{\mathrm{vec}\Big(\sum_{i\in\mathcal{I}_k} \boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\Big) + \mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Bigg\} \\
&- 2^{-1}\sum_{k=1}^K \big(N_k-1\big)\mathbf{D}^\top\mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\big) - 2^{-1}\sum_{k=1}^K \mathbf{D}^\top\mathrm{vec}\Big\{N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}.
\end{aligned}
$$

Here, $\widetilde{\boldsymbol{\alpha}}_m = \alpha_m^{-1}\boldsymbol{e}_m \in \mathbb{R}^M$ for $1 \le m \le M-1$, $\widetilde{\boldsymbol{\alpha}}_M = -(1-\sum_m^{M-1}\alpha_m)^{-1}\mathbf{1}$, where $\boldsymbol{e}_m$ denotes the mth column of the identity matrix $\mathbf{I}$. Therefore, we computed the result for the first-order derivative of $\mathcal{L}(\boldsymbol{\theta})$.

Step 2. Consequently, we should have a gradient condition as $\dot{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}) = \mathbf{0}$. This suggests a set of estimation equations for parameters (i.e., α_ms, $\boldsymbol{\nu}_{m'}$s, $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$), respectively. By setting $\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \mathbf{0}$, which can be rewritten as

$$\sum_{k=1}^{K}\sum_{m=1}^{M-1} \pi_{(\mathcal{I}_k)}^{m} \widetilde{\boldsymbol{\alpha}}_m = -\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{M} \widetilde{\boldsymbol{\alpha}}_M.$$

For each α_m with $1 \leq m \leq M-1$, we have $\alpha_m = (\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m})(\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{M} \alpha_M^{-1})^{-1}$. Since $\sum_{m=1}^{M} \alpha_m = 1$, we have $\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{M} \alpha_M^{-1} = K$. Therefore, we can obtain $\widehat{\alpha}_{m,\mathrm{MLE}} = \mathcal{F}_{\alpha_m}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}})$ for $1 \leq m \leq M-1$ as $\mathcal{F}_{\alpha_m}(\boldsymbol{\theta}) = K^{-1}\left(\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m}\right)$. Let $\dot{\mathcal{L}}_{\boldsymbol{\nu}}(\boldsymbol{\theta}) = \mathbf{0}$. Then, we have $\widehat{\boldsymbol{\nu}}_{m',\mathrm{MLE}} = \mathcal{F}_{\boldsymbol{\nu}_{m'}}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}})$ with $1 \leq m' \leq M$, where

$$\mathcal{F}_{\boldsymbol{\nu}_{m'}}(\boldsymbol{\theta}) = \left\{\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m'} \left(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\right)^{-1}\right\}^{-1} \left\{\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m'} \left(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\right)^{-1} \overline{\boldsymbol{X}}_k\right\}.$$

Next, we compute $\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) = \mathbf{0}$ as

$$2^{-1}\mathbf{D}^{\top}\mathrm{vec}\Big[\sum_{k=1}^{K}\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top} - \sum_{k=1}^{K} (N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\Big] = \mathbf{0}.$$

Subsequently, the mapping function can be obtained by calculating $\sum_{k=1}^{K}(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1} = \sum_{k=1}^{K}\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}$. Since the matrices $N_k^{-1}\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are invertible, we have $(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_k^{*}\boldsymbol{\Omega}^{-1}$, $\boldsymbol{\Sigma}_k^{*} = \boldsymbol{\Omega}(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}N_k^{-1}\boldsymbol{\Sigma}$, and the equation as $K\boldsymbol{\Omega} = \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{*} + \sum_{k=1}^{K}\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \boldsymbol{\Omega}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\boldsymbol{\Omega}$. Thus, we have $\widehat{\boldsymbol{\Omega}}_{\mathrm{MLE}} = \mathcal{F}_{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}})$ as

$$\begin{aligned}\mathcal{F}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) =& K^{-1}\Bigg\{\sum_{k=1}^{K}\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \boldsymbol{\Omega}\left(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\right)^{-1}\left(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\right) \\ & \times \left(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\right)^{\top}\left(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\right)^{-1}\boldsymbol{\Omega} + \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{*}\Bigg\}.\end{aligned}$$

Then, the condition $\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) = \mathbf{0}$ can be written as

$$\begin{aligned}\mathbf{0} =& 2^{-1}\mathbf{D}^{\top}\mathrm{vec}\Bigg[\sum_{k=1}^{K}\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m}\Big(\sum_{i\in\mathcal{I}_k} \boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^{\top} + N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\Big) \\ & - \sum_{k=1}^{K}(N_k - 1)\boldsymbol{\Sigma}^{-1} - \sum_{k=1}^{K} N_k^{-1}\left(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega}\right)^{-1}\Bigg],\end{aligned}$$

which is same to calculate the following equation

$$\begin{aligned}\big(N-K\big)\boldsymbol{\Sigma} =&\boldsymbol{\Sigma}\Big[\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\Big(\sum_{i\in\mathcal{I}_k}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top+N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\\&-\sum_{k=1}^{K}N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big]\boldsymbol{\Sigma}.\end{aligned}$$

Therefore, we have $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MLE}}=\mathcal{F}_{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}})$ as

$$\begin{aligned}\mathcal{F}_{\boldsymbol{\Sigma}}\left(\boldsymbol{\theta}\right)=&\left(N-K\right)^{-1}\left(\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\Big[\sum_{i\in\mathcal{I}_k}\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k\right)\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k\right)^\top\right.\\&+N_k^{-1}\Big\{\boldsymbol{\Sigma}\left(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\right)^{-1}\left(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m\right)\left(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m\right)^\top\left(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\right)^{-1}\boldsymbol{\Sigma}\\&\left.-\boldsymbol{\Sigma}\left(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\right)^{-1}\boldsymbol{\Sigma}\Big\}\Big]\right).\end{aligned}$$

By this point, we have derived the analytical form of mapping functions for iterative parameter estimation as well as the posterior probabilities. This completes the proof of the mapping functions in equation (3).

Step 3. Lastly, the expectation and variance of the first order derivative of $\mathcal{L}(\boldsymbol{\theta})$ can be obtained. To this end, we start with verifying that $E\{\dot{\mathcal{L}}(\boldsymbol{\theta})\}=\mathbf{0}$. For $\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})$, we can compute $E\{\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})\}=\sum_{k=1}^{K}\sum_{m=1}^{M}E\{\pi_{(\mathcal{I}_k)}^{m}\}\widetilde{\boldsymbol{\alpha}}_m=\sum_{k=1}^{K}\sum_{m=1}^{M}\alpha_m\widetilde{\boldsymbol{\alpha}}_m=\mathbf{0}$. Then, we should calculate $E\{\dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta})\}=(E\{\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\boldsymbol{b}_{k1}^\top\},\ldots,E\{\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\boldsymbol{b}_{kM}^\top\})^\top$. It can be obtained that $E\{\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\boldsymbol{b}_{km}^\top\}=\sum_{k=1}^{K}E[E\{\mathbf{1}(Z_k=m)\boldsymbol{b}_{km}^\top|\mathbf{X}_{(\mathcal{I}_k)},\boldsymbol{Y}_{(\mathcal{I}_k)}\}]=\sum_{k=1}^{K}\alpha_mE(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m|Z_k=m)(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}$. Since $E(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m|Z_k=m)=\int E(\overline{\boldsymbol{X}}_k-\boldsymbol{\mu}_{km}+\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m|Z_k=m,\boldsymbol{\mu}_{km})\phi_{\boldsymbol{\nu}_m,\boldsymbol{\Omega}}(\boldsymbol{\mu}_{km})d\boldsymbol{\mu}_{km}=\mathbf{0}$, we have $E\{\dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta})\}=\mathbf{0}$. The technical details for proving both $E\{\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta})\}$ and $E\{\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})\}$ are similar. Comparatively speaking, the proof of $E\{\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})\}$ is more complicated. Thus, we provide the proof details for $E\{\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})\}$ only. The proof for $E\{\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta})\}$ is omitted to save space. Recall that

$$\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) =2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\mathbf{D}^\top\left\{\mathrm{vec}\Big(\sum_{i\in\mathcal{I}_k}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\Big)+\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\right\}$$

$$- 2^{-1} \sum_{k=1}^{K} \big(N_k - 1\big) \mathbf{D}^\top \mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\big) - 2^{-1} \sum_{k=1}^{K} \mathbf{D}^\top \mathrm{vec}\Big\{ N_k^{-1} \big(N_k^{-1} \boldsymbol{\Sigma} + \boldsymbol{\Omega}\big)^{-1} \Big\}.$$

We can similarly prove the expectation of the first part of $\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$ as $2^{-1} \sum_{k=1}^{K} \sum_{m=1}^{M} \alpha_m$ $\mathbf{D}^\top E\{\mathrm{vec}(\sum_{i \in \mathcal{I}_k} \boldsymbol{c}_{ik} \boldsymbol{c}_{ik}^\top) + \mathrm{vec}(N_k^{-1} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^\top) | Z_k = m\}$. To address this, we need to compute $E\{\mathrm{vec}(\sum_{i \in \mathcal{I}_k} \boldsymbol{c}_{ik} \boldsymbol{c}_{ik}^\top) | Z_k = m\}$ and $E\{\mathrm{vec}(N_k^{-1} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^\top) | Z_k = m\}$ separately. We have $E\{\mathrm{vec}(\sum_{i \in \mathcal{I}_k} \boldsymbol{c}_{ik} \boldsymbol{c}_{ik}^\top) | Z_k = m\} = \sum_{i \in \mathcal{I}_k} \mathrm{vec}[\boldsymbol{\Sigma}^{-1} E\{(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)^\top | Z_k = m\} \boldsymbol{\Sigma}^{-1}] = (N_k - 1) \mathrm{vec}(\boldsymbol{\Sigma}^{-1})$. We then compute $E\{\mathrm{vec}(N_k^{-1} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^\top) | Z_k = m\} = \mathrm{vec}[N_k^{-1} (N_k^{-1} \boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1} E\{(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)^\top | Z_k = m\} (N_k^{-1} \boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}] = \mathrm{vec}\{N_k^{-1} (N_k^{-1} \boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\}$. Therefore, based on the above results, we can verify that $E\{\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})\} = \mathbf{0}$.

Next, we compute the covariance of $\dot{\mathcal{L}}(\boldsymbol{\theta})$. Since $E\{\dot{\mathcal{L}}(\boldsymbol{\theta})\} = \mathbf{0}$, we have $\mathrm{cov}\{\dot{\mathcal{L}}(\boldsymbol{\theta})\} = E\{\dot{\mathcal{L}}(\boldsymbol{\theta}) \dot{\mathcal{L}}^\top(\boldsymbol{\theta})\} = \boldsymbol{\Gamma} \in \mathbb{R}^{q \times q}$, where $\boldsymbol{\Gamma}_{11} = E\{\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\alpha}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{(M-1) \times (M-1)}$, $\boldsymbol{\Gamma}_{22} = (\Gamma_{22,m_1 m_2}) = E\{\dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\nu}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{Mp \times Mp}$, with $\boldsymbol{\Gamma}_{22,m_1 m_2} \in \mathbb{R}^{p \times p}$ for $1 \le m_1,\, m_2 \le M$, $\boldsymbol{\Gamma}_{33} = E\{\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Omega}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{\{p(p+1)/2\} \times \{p(p+1)/2\}}$, and $\boldsymbol{\Gamma}_{44} = E\{\dot{\mathcal{L}}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Sigma}}^\top(\boldsymbol{\theta})\} \times \{p(p+1)/2\}$. The off-diagonal block matrices are given by $\boldsymbol{\Gamma}_{12} = \boldsymbol{\Gamma}_{21}^\top = (\boldsymbol{\Gamma}_{12,1}, \ldots, \boldsymbol{\Gamma}_{12,M}) = E\{\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\nu}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{(M-1) \times Mp}$, with $\boldsymbol{\Gamma}_{12,m} \in \mathbb{R}^{(M-1) \times p}$ and $1 \le m \le M$, $\boldsymbol{\Gamma}_{13} = \boldsymbol{\Gamma}_{31}^\top = E\{\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Omega}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{(M-1) \times \{p(p+1)/2\}}$, $\boldsymbol{\Gamma}_{14} = \boldsymbol{\Gamma}_{41}^\top = E\{\dot{\mathcal{L}}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Sigma}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{(M-1) \times \{p(p+1)/2\}}$, and $\boldsymbol{\Gamma}_{23} = \boldsymbol{\Gamma}_{32}^\top = (\boldsymbol{\Gamma}_{23,1}^\top, \ldots, \boldsymbol{\Gamma}_{23,M}^\top)^\top = E\{\dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Omega}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{Mp \times \{p(p+1)/2\}}$, with $\boldsymbol{\Gamma}_{23,m} \in \mathbb{R}^{p \times \{p(p+1)/2\}}$ and $1 \le m \le M$, and $\boldsymbol{\Gamma}_{24} = \boldsymbol{\Gamma}_{42}^\top = (\boldsymbol{\Gamma}_{24,1}^\top, \ldots, \boldsymbol{\Gamma}_{24,M}^\top)^\top = E\{\dot{\mathcal{L}}_{\nu}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Sigma}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{Mp \times \{p(p+1)/2\}}$, with $\boldsymbol{\Gamma}_{24,m} \in \mathbb{R}^{p \times \{p(p+1)/2\}}$, and $\boldsymbol{\Gamma}_{34} = \boldsymbol{\Gamma}_{43}^\top = E\{\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) \dot{\mathcal{L}}_{\boldsymbol{\Sigma}}^\top(\boldsymbol{\theta})\} \in \mathbb{R}^{\{p(p+1)/2\}} \times \mathbb{R}^{\{p(p+1)/2\}}$. We first compute $\boldsymbol{\Gamma}_{11} = E\{(\sum_{k=1}^{K} \sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \widetilde{\boldsymbol{\alpha}}_m)(\sum_{k'=1}^{K} \sum_{m'=1}^{M} \pi_{(\mathcal{I}_{k'})}^{m'} \widetilde{\boldsymbol{\alpha}}_{m'})\}$. Since individuals from different classes are independent, it can be proved that $\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \widetilde{\boldsymbol{\alpha}}_m$ is independent with $\sum_{m'=1}^{M} \pi_{(\mathcal{I}_{k'})}^{m'} \widetilde{\boldsymbol{\alpha}}_{m'}$ for $k \neq k'$. Then, we can obtain

$$\boldsymbol{\Gamma}_{11} = \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{m'=1}^{M} E(\pi_{(\mathcal{I}_k)}^{m} \pi_{(\mathcal{I}_k)}^{m'}) \widetilde{\boldsymbol{\alpha}}_m \widetilde{\boldsymbol{\alpha}}_{m'}^\top.$$

Then, we calculate $\boldsymbol{\Gamma}_{22}$ by decomposing it as a set of block matrices as $\boldsymbol{\Gamma}_{22,m_1 m_2}$s for $1 \le m_1, m_2 \le M$. Accordingly, we compute $\boldsymbol{\Gamma}_{22,m_1 m_2} = \sum_{k=1}^{K} E(\pi_{(\mathcal{I}_k)}^{m_1} \pi_{(\mathcal{I}_k)}^{m_2} \boldsymbol{b}_{km_1} \boldsymbol{b}_{km_2}^\top)$. Next, we study $\boldsymbol{\Gamma}_{33} = 4^{-1} \sum_{k=1}^{K} E\Big\{ \big(\sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \mathbf{D}^\top \big[\mathrm{vec}(\boldsymbol{b}_{km} \boldsymbol{b}_{km}^\top) - \mathrm{vec}\{(N_k^{-1} \boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\} \big] \big)$

$\big(\sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\mathbf{D}^\top\big[\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top)-\mathrm{vec}\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\big]\big)^\top\Big\}$. We then obtain that

$$
\begin{aligned}
&E\Bigg\{\Bigg(\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\mathbf{D}^\top\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\Bigg)\Bigg[\mathbf{D}^\top\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\Bigg]^\top\Bigg\}\\
=&\sum_{m=1}^{M}\alpha_m\mathbf{D}^\top E\Big\{\mathrm{vec}\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)|Z_k=m\Big\}\Bigg[\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\Bigg]^\top\mathbf{D}\\
=&\mathbf{D}^\top\Bigg[\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\Bigg]\Bigg[\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\Bigg]^\top\mathbf{D}.
\end{aligned}
$$

We then obtain $\boldsymbol{\Gamma}_{33}=4^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbf{D}^\top E[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top)\}^\top]\mathbf{D}-4^{-1}\sum_{k=1}^{K}\mathbf{D}^\top\mathrm{vec}\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\big[\mathrm{vec}\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\big]^\top\mathbf{D}$. By the similar arguments, $\boldsymbol{\Gamma}_{44}$ can be given as

$$
\begin{aligned}
\boldsymbol{\Gamma}_{44}=&-4^{-1}\sum_{k=1}^{K}\mathbf{D}^\top\Bigg[N_k^{-2}\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}^\top\Bigg]\mathbf{D}\\
&+2^{-1}\sum_{k=1}^{K}\mathbf{D}^\top\Big\{(N_k-1)\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\big)\Big\}\mathbf{D}+4^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbf{D}^\top E\Bigg[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\\
&\times\Big\{N_k^{-1}\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}\Big\{N_k^{-1}\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top\Bigg]\mathbf{D}.
\end{aligned}
$$

Next, we focus on computing the off-diagonal matrices. By decomposing $\boldsymbol{\Gamma}_{12}=(\boldsymbol{\Gamma}_{12,1},\dots,\boldsymbol{\Gamma}_{12,M})$, we can compute $\boldsymbol{\Gamma}_{12,m}=\sum_{k=1}^{K}E\{(\sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\widetilde{\boldsymbol{\alpha}}_{m'})(\pi_{(\mathcal{I}_k)}^{m}\boldsymbol{b}_{km}^\top)\}=\sum_{k=1}^{K}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_{m'}E(\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\boldsymbol{b}_{km}^\top)$. Similar calculation leads to $\boldsymbol{\Gamma}_{13}=E\Big\{(\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\widetilde{\boldsymbol{\alpha}}_m)\big(2^{-1}\sum_{k'=1}^{K}\sum_{m'=1}^{M}\pi_{(\mathcal{I}_{k'})}^{m'}\mathbf{D}^\top[\mathrm{vec}(\boldsymbol{b}_{k'm'}\boldsymbol{b}_{k'm'}^\top)-\mathrm{vec}\{(N_{k'}^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}]\big)^\top\Big\}$, which can be verified as

$$
\begin{aligned}
\boldsymbol{\Gamma}_{13}&=2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_m E\Bigg[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\Big\{\mathrm{vec}\big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\big)\Big\}^\top\Bigg]\mathbf{D},\\
\boldsymbol{\Gamma}_{14}&=2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_m E\Bigg[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\Big\{\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top\Bigg]\mathbf{D}.
\end{aligned}
$$

Write $\boldsymbol{\Gamma}_{23}=(\boldsymbol{\Gamma}_{23,1}^\top,\dots,\boldsymbol{\Gamma}_{23,M}^\top)^\top$, where $\boldsymbol{\Gamma}_{23,m}=\mathrm{cov}(\dot{\mathcal{L}}_{\boldsymbol{\nu}_m}(\boldsymbol{\theta}),\dot{\mathcal{L}}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}))$. Since we have proved $E\{\sum_{k=1}^{K} \pi^{m}_{(\mathcal{I}_k)} \boldsymbol{b}_{km}^\top\} = \mathbf{0}$, it can be verified that

$$\mathbf{\Gamma}_{23,m} = 2^{-1} \sum_{k=1}^{K} \sum_{m'=1}^{M} E\left[\pi^{m}_{(\mathcal{I}_k)} \pi^{m'}_{(\mathcal{I}_k)} \boldsymbol{b}_{km} \left\{\mathrm{vec}\left(\boldsymbol{b}_{km'} \boldsymbol{b}_{km'}^\top\right)\right\}^\top\right] \mathbf{D}.$$

By similar technique, we have $\mathbf{\Gamma}_{24} = (\mathbf{\Gamma}_{24,1}^\top, \dots, \mathbf{\Gamma}_{24,M}^\top)^\top$, where

$$\mathbf{\Gamma}_{24,m} = 2^{-1} \sum_{k=1}^{K} \sum_{m'=1}^{M} E\left[\pi^{m}_{(\mathcal{I}_k)} \pi^{m'}_{(\mathcal{I}_k)} \boldsymbol{b}_{km} \left\{\mathrm{vec}\left(N_k^{-1} \boldsymbol{b}_{km'} \boldsymbol{b}_{km'}^\top\right)\right\}^\top\right] \mathbf{D}.$$

Subsequently, we can also compute $\mathbf{\Gamma}_{34}$ as

$$\begin{aligned}
\mathbf{\Gamma}_{34} = &- 4^{-1} \sum_{k=1}^{K} \mathbf{D}^\top \left[N_k^{-1} \mathrm{vec}\left\{\left(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\right)^{-1}\right\} \mathrm{vec}\left\{\left(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\right)^{-1}\right\}^\top\right] \mathbf{D} \\
&+ 4^{-1} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{m'=1}^{M} \mathbf{D}^\top E\left[\pi^{m}_{(\mathcal{I}_k)} \pi^{m'}_{(\mathcal{I}_k)} \left\{\mathrm{vec}\left(\boldsymbol{b}_{km} \boldsymbol{b}_{km}^\top\right)\right\} \left\{N_k^{-1} \mathrm{vec}\left(\boldsymbol{b}_{km'} \boldsymbol{b}_{km'}^\top\right)\right\}^\top\right] \mathbf{D}.
\end{aligned}$$

In conclusion, we compute the first derivative of the log-likelihood function and then obtain the mapping function needed in the EM algorithm. The expectation and covariance of the first derivation are also calculated for the following theorem.

Appendix A.3: The Derivation of $\ddot{\mathcal{L}}(\boldsymbol{\theta})$ and its Expectation

We consider computing the second derivatives of $\mathcal{L}(\boldsymbol{\theta})$, followed by a subsequent calculation of their expectations.

Step 1. We begin with computing $\ddot{\mathcal{L}}(\boldsymbol{\theta})$ by derivative. By decomposing $\mathcal{L}(\boldsymbol{\theta})$ as (A.4), the second derivative of $\mathcal{L}(\boldsymbol{\theta})$ can be given by

$$d^2\mathcal{L}(\boldsymbol{\theta}) = d^2\mathcal{L}_1(\boldsymbol{\theta}) + d^2\mathcal{L}_2(\boldsymbol{\theta}) + d^2\mathcal{L}_3(\boldsymbol{\theta}) + d^2\mathcal{L}_4(\boldsymbol{\theta}),$$

whose terms can be computed separately. Recall that $d\mathcal{L}_1(\boldsymbol{\theta}) = -2^{-1}N\mathrm{tr}(\mathbf{\Sigma}^{-1}d\mathbf{\Sigma})$. We then have $d^2\mathcal{L}_1(\boldsymbol{\theta}) = -2^{-1}N\mathrm{tr}\{d(\mathbf{\Sigma}^{-1})d\mathbf{\Sigma}\} = 2^{-1}N\mathrm{tr}\{\mathbf{\Sigma}^{-1}d(\mathbf{\Sigma})\mathbf{\Sigma}^{-1}d\mathbf{\Sigma}\}$. By relevant simple matrix calculation, we can obtain

$$d^2\mathcal{L}_1(\boldsymbol{\theta}) = 2^{-1}N\Big\{d\text{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top \mathbf{D}^\top\Big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\Big)\mathbf{D}\Big\{d\text{vech}\big(\boldsymbol{\Sigma}\big)\Big\}. \tag{A.5}$$

Since $d\mathcal{L}_2(\boldsymbol{\theta}) = -2^{-1}K\text{tr}(\boldsymbol{\Omega}^{-1}d\boldsymbol{\Omega})$, we similarly have

$$d^2\mathcal{L}_2(\boldsymbol{\theta}) = 2^{-1}K\Big\{d\text{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top \mathbf{D}^\top\Big(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\Big)\mathbf{D}\Big\{d\text{vech}\big(\boldsymbol{\Omega}\big)\Big\}. \tag{A.6}$$

Next, we need to calculate $d^2\mathcal{L}_3(\boldsymbol{\theta})$, where $d\mathcal{L}_3(\boldsymbol{\theta}) = 2^{-1}\sum_{k=1}^K[\text{tr}\{\boldsymbol{\Omega}^{-1} - (N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}d\boldsymbol{\Omega}\} + \text{tr}\{\boldsymbol{\Sigma}^{-1} - N_k^{-1}(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\Sigma}\}]$. We have $d^2\mathcal{L}_3(\boldsymbol{\theta}) = -2^{-1}\sum_{k=1}^K[\text{tr}\{\boldsymbol{\Omega}^{-1} d(\boldsymbol{\Omega})\boldsymbol{\Omega}^{-1}d(\boldsymbol{\Omega}) - (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\Omega}\} + \text{tr}\{\boldsymbol{\Sigma}^{-1}d(\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}d(\boldsymbol{\Sigma}) - N_k^{-1}(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\Sigma}\}]$. Thus, it can be rewritten as

$$\begin{aligned} d^2\mathcal{L}_3(\boldsymbol{\theta}) =& -2^{-1}\sum_{k=1}^K\bigg[\Big\{d\text{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top\mathbf{D}^\top\Big\{\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1} - \big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1} \\ &\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Omega}\big) + \Big\{d\text{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top\mathbf{D}^\top\Big\{\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1} \\ &- N_k^{-2}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Sigma}\big) - \Big\{d\text{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top\mathbf{D}^\top \\ &\times\Big\{N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}(\boldsymbol{\Omega}) \\ &- \Big\{d\text{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top\mathbf{D}^\top\Big\{N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Sigma}\big)\bigg]. \end{aligned} \tag{A.7}$$

As $\mathcal{L}_4(\boldsymbol{\theta}) = \sum_{k=1}^K\log\Big(\sum_{m=1}^M\tau_{mk}\Big)$, the calculation of $d^2\mathcal{L}_4(\boldsymbol{\theta})$ can be proved as

$$\begin{aligned} d^2\mathcal{L}_4(\boldsymbol{\theta}) =& \sum_{k=1}^K d^2\log\left(\sum_{m=1}^M\tau_{mk}\right) = \sum_{k=1}^K\left\{\frac{\sum_{m=1}^M d^2\tau_{mk}}{\sum_{m'=1}^M\tau_{m'k}} - \left(\frac{\sum_{m=1}^M d\tau_{mk}}{\sum_{m'=1}^M\tau_{m'k}}\right)^2\right\} \\ =& \sum_{k=1}^K\left[\sum_{m=1}^M\pi_{(\mathcal{I}_k)}^m\Big\{d^2\log\tau_{mk} + \big(d\log\tau_{mk}\big)^2\Big\} - \Big(\sum_{m=1}^M\pi_{(\mathcal{I}_k)}^m d\log\tau_{mk}\Big)^2\right], \end{aligned} \tag{A.8}$$

To this end, we need to calculate the derivative of each term of $d\log\tau_{mk}$, where $d\log\tau_{mk} = \alpha_m^{-1}d\alpha_m + \boldsymbol{b}_{km}^\top d\boldsymbol{\nu}_m + 2^{-1}\{\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}d\text{vech}(\boldsymbol{\Omega}) + 2^{-1}\{\text{vec}(\sum_{i\in\mathcal{I}_k}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top) + N_k^{-1}\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}d\text{vech}(\boldsymbol{\Sigma})$. The first term can be obtained simply as $d(\alpha_m^{-1})d\alpha_m =$

$-d(\alpha_m)\alpha_m^{-2}d(\alpha_m)$. The second term $d(\boldsymbol{b}_{km})^\top d\boldsymbol{\nu}_m = \{d(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)\}^\top (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\nu}_m + (\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)^\top d\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}d\boldsymbol{\nu}_m = -(d\boldsymbol{\nu}_m)^\top (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\nu}_m - \boldsymbol{b}_{km}^\top d(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\nu}_m = -(d\boldsymbol{\nu}_m)^\top (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}d\boldsymbol{\nu}_m - \{d\text{vech}(\boldsymbol{\Omega})\}^\top \mathbf{D}^\top \{\boldsymbol{b}_{km}\otimes (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}d\boldsymbol{\nu}_m - \{d\text{vech}(\boldsymbol{\Sigma})\}^\top \mathbf{D}^\top\{\boldsymbol{b}_{km}\otimes N_k^{-1}(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}d\boldsymbol{\nu}_m$. Next, the third term of $d\log\tau_{mk}$, $2^{-1}\{d\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top \mathbf{D}d\text{vech}(\boldsymbol{\Omega})$ can be verified as

$$
\begin{aligned}
&2^{-1}\Big\{d\text{vec}\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\Big\}^\top \mathbf{D}d\text{vech}\big(\boldsymbol{\Omega}\big) \\
=&-\bigg[\text{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}d\boldsymbol{\nu}_m\boldsymbol{b}_{km}^\top + \big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}d\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big\}\bigg]^\top \\
&\times \mathbf{D}d\text{vech}\big(\boldsymbol{\Omega}\big) \\
=&-\big(d\boldsymbol{\nu}_m\big)^\top\Big\{\boldsymbol{b}_{km}^\top\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Omega}\big) \\
&-\Big\{d\text{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top\mathbf{D}^\top\Big\{\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Omega}\big) \\
&-\Big\{d\text{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top\mathbf{D}^\top\Big\{\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\otimes N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Omega}\big).
\end{aligned}
$$

Lastly, the fourth term can be computed as follows

$$
\begin{aligned}
&2^{-1}\bigg[\sum_{i\in\mathcal{I}_k}\text{vec}\Big\{d\big(\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\big)\Big\} + N_k^{-1}\text{vec}\Big\{d\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\Big\}\bigg]^\top\mathbf{D}d\text{vech}\big(\boldsymbol{\Sigma}\big) \\
=&-N_k^{-1}\big(d\boldsymbol{\nu}_m\big)^\top\Big\{\boldsymbol{b}_{km}^\top\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Sigma}\big) - N_k^{-1}\Big\{d\text{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top\mathbf{D}^\top \\
&\times\Big\{\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Sigma}\big) - \Big\{d\text{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top\mathbf{D}^\top \\
&\times\Big\{\sum_{i\in\mathcal{I}_k}\big(\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\big)\otimes\boldsymbol{\Sigma}^{-1} + \big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\otimes N_k^{-2}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}d\text{vech}\big(\boldsymbol{\Sigma}\big).
\end{aligned}
$$

Combing above results, we then have

$$
d^2\log\tau_{mk} = -\begin{bmatrix} d\boldsymbol{\alpha} \\ d\boldsymbol{\nu} \\ d\text{vech}(\boldsymbol{\Omega}) \\ d\text{vech}(\boldsymbol{\Sigma}) \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{\tau}_{mk,11}^{d2} & \boldsymbol{\tau}_{mk,12}^{d2} & \boldsymbol{\tau}_{mk,13}^{d2} & \boldsymbol{\tau}_{mk,14}^{d2} \\ \boldsymbol{\tau}_{mk,21}^{d2} & \boldsymbol{\tau}_{mk,22}^{d2} & \boldsymbol{\tau}_{mk,23}^{d2} & \boldsymbol{\tau}_{mk,24}^{d2} \\ \boldsymbol{\tau}_{mk,31}^{d2} & \boldsymbol{\tau}_{mk,32}^{d2} & \boldsymbol{\tau}_{mk,33}^{d2} & \boldsymbol{\tau}_{mk,34}^{d2} \\ \boldsymbol{\tau}_{mk,41}^{d2} & \boldsymbol{\tau}_{mk,42}^{d2} & \boldsymbol{\tau}_{mk,43}^{d2} & \boldsymbol{\tau}_{mk,44}^{d2} \end{bmatrix} \begin{bmatrix} d\boldsymbol{\alpha} \\ d\boldsymbol{\nu} \\ d\text{vech}(\boldsymbol{\Omega}) \\ d\text{vech}(\boldsymbol{\Sigma}) \end{bmatrix},
$$

where the block matrices located on the diagonal are $\boldsymbol{\tau}^{d2}_{mk,11} = \partial^2 \log \tau_{mk}/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top = \widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_m^\top \in \mathbb{R}^{(M-1)\times(M-1)}$, $\boldsymbol{\tau}^{d2}_{mk,22} = (\boldsymbol{\tau}^{d2}_{mk,22,m_1m_2}) = \partial^2 \log \tau_{mk}/\partial\boldsymbol{\alpha}\partial\boldsymbol{\nu}^\top \in \mathbb{R}^{Mp\times Mp}$ with $\boldsymbol{\tau}^{d2}_{mk,22,mm} = (N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1} \in \mathbb{R}^{p\times p}$, and other elements equal to $\mathbf{0}$, $\boldsymbol{\tau}^{d2}_{mk,33} = \partial^2 \log \tau_{mk}/ \partial\text{vech}(\boldsymbol{\Omega})\partial\text{vech}(\boldsymbol{\Omega})^\top = \mathbf{D}^\top\{(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top) \otimes (N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\}\mathbf{D} \in \mathbb{R}^{\{p(p+1)/2\}\times\{p(p+1)/2\}}$, $\boldsymbol{\tau}^{d2}_{mk,44} = \partial^2 \log \tau_{mk}/\partial\text{vech}(\boldsymbol{\Sigma})\partial\text{vech}(\boldsymbol{\Sigma})^\top = \mathbf{D}^\top\{\sum_{i\in\mathcal{I}_k}(\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top) \otimes \boldsymbol{\Sigma}^{-1} + (\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top) \otimes N_k^{-2}(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\}\mathbf{D} \in \mathbb{R}^{\{p(p+1)/2\}\times\{p(p+1)/2\}}$, and the off-diagonal block matrices can be calculated as $\boldsymbol{\tau}^{d2}_{mk,12} = \boldsymbol{\tau}^{d2\top}_{mk,21} = \partial^2 \log \tau_{mk}/\partial\boldsymbol{\alpha}\partial\boldsymbol{\nu}^\top = \mathbf{0} \in \mathbb{R}^{(M-1)\times Mp}$, $\boldsymbol{\tau}^{d2}_{mk,13} = \boldsymbol{\tau}^{d2\top}_{mk,31} = \partial^2 \log \tau_{mk}/\partial\boldsymbol{\alpha}\partial\text{vech}(\boldsymbol{\Omega})^\top = \mathbf{0} \in \mathbb{R}^{(M-1)\times\{p(p+1)/2\}}$, $\boldsymbol{\tau}^{d2}_{mk,14} = \boldsymbol{\tau}^{d2\top}_{mk,41} = \partial^2 \log \tau_{mk}/ \partial\boldsymbol{\alpha}\partial\text{vech}(\boldsymbol{\Sigma})^\top = \mathbf{0} \in \mathbb{R}^{(M-1)\times\{p(p+1)/2\}}$, $\boldsymbol{\tau}^{d2}_{mk,23} = \boldsymbol{\tau}^{d2\top}_{mk,32} = \partial^2 \log \tau_{mk}/\partial\boldsymbol{\nu}\partial\text{vech}(\boldsymbol{\Omega})^\top = \{\boldsymbol{b}_{km}^\top\otimes(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\mathbf{D} \in \mathbb{R}^{Mp\times\{p(p+1)/2\}}$, $\boldsymbol{\tau}^{d2}_{mk,24} = \boldsymbol{\tau}^{d2\top}_{mk,42} = \partial^2 \log \tau_{mk}/\partial\boldsymbol{\nu}\partial\text{vech}(\boldsymbol{\Sigma})^\top = \{\boldsymbol{b}_{km}^\top\otimes N_k^{-1}(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\mathbf{D} \in \mathbb{R}^{Mp\times\{p(p+1)/2\}}$, and $\boldsymbol{\tau}^{d2}_{mk,34} = \boldsymbol{\tau}^{d2\top}_{mk,43} = \partial^2 \log \tau_{mk}/\partial\text{vech}(\boldsymbol{\Omega}) \partial\text{vech}(\boldsymbol{\Sigma})^\top = \mathbf{D}^\top\{(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top) \otimes N_k^{-1}(N_k^{-1}\boldsymbol{\Sigma} + \boldsymbol{\Omega})^{-1}\}\mathbf{D} \in \mathbb{R}^{\{p(p+1)/2\}\times\{p(p+1)/2\}}$.

Through matrix multiplication, we can obtain that

$$\Big(d\log\tau_{mk}\Big)^2 = \begin{bmatrix} d\boldsymbol{\alpha} \\ d\boldsymbol{\nu} \\ d\text{vech}(\boldsymbol{\Omega}) \\ d\text{vech}(\boldsymbol{\Sigma}) \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{\tau}^{d1}_{mk,11} & \boldsymbol{\tau}^{d1}_{mk,12} & \boldsymbol{\tau}^{d1}_{mk,13} & \boldsymbol{\tau}^{d1}_{mk,14} \\ \boldsymbol{\tau}^{d1}_{mk,21} & \boldsymbol{\tau}^{d1}_{mk,22} & \boldsymbol{\tau}^{d1}_{mk,23} & \boldsymbol{\tau}^{d1}_{mk,24} \\ \boldsymbol{\tau}^{d1}_{mk,31} & \boldsymbol{\tau}^{d1}_{mk,32} & \boldsymbol{\tau}^{d1}_{mk,33} & \boldsymbol{\tau}^{d1}_{mk,34} \\ \boldsymbol{\tau}^{d1}_{mk,41} & \boldsymbol{\tau}^{d1}_{mk,42} & \boldsymbol{\tau}^{d1}_{mk,43} & \boldsymbol{\tau}^{d1}_{mk,44} \end{bmatrix} \begin{bmatrix} d\boldsymbol{\alpha} \\ d\boldsymbol{\nu} \\ d\text{vech}(\boldsymbol{\Omega}) \\ d\text{vech}(\boldsymbol{\Sigma}) \end{bmatrix},$$

where $\boldsymbol{\tau}^{d1}_{mk,11} = \widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_m^\top$, $\boldsymbol{\tau}^{d1}_{mk,22} = (\boldsymbol{\tau}^{d2}_{mk,22,m_1m_2})$ with $\boldsymbol{\tau}^{d1}_{mk,22,mm} = \boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top$, and other elements equal to $\mathbf{0}$, $\boldsymbol{\tau}^{d1}_{mk,33} = 4^{-1}\mathbf{D}^\top\{\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, $\boldsymbol{\tau}^{d1}_{mk,44} = 4^{-1}\mathbf{D}^\top\{ \text{vec}(\sum_{i\in\mathcal{I}_K}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top) + N_k^{-1}\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\text{vec}(\sum_{i\in\mathcal{I}_K}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top) + N_k^{-1}\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, and the off-diagonal block matrices are $\boldsymbol{\tau}^{d1}_{mk,12} = (\boldsymbol{\tau}^{d1}_{mk,12,1}, \dots, \boldsymbol{\tau}^{d1}_{mk,12,M}) = \boldsymbol{\tau}^{d1\top}_{mk,21}$ with $\boldsymbol{\tau}^{d1}_{mk,12,m} = \widetilde{\boldsymbol{\alpha}}_m\boldsymbol{b}_{km}^\top \in \mathbb{R}^{(M-1)\times p}$, $\boldsymbol{\tau}^{d1}_{mk,13} = \boldsymbol{\tau}^{d1\top}_{mk,31} = 2^{-1}\widetilde{\boldsymbol{\alpha}}_m\{\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, $\boldsymbol{\tau}^{d1}_{mk,14} = \boldsymbol{\tau}^{d1\top}_{mk,41} = 2^{-1}\widetilde{\boldsymbol{\alpha}}_m\{\text{vec}(\sum_{i\in\mathcal{I}_K}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top)+N_k^{-1}\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, $\boldsymbol{\tau}^{d1}_{mk,23} = \boldsymbol{\tau}^{d1\top}_{mk,32} = (\boldsymbol{\tau}^{d1\top}_{mk,23,1}, \dots, \boldsymbol{\tau}^{d1\top}_{mk,23,M})^\top$, with $\boldsymbol{\tau}^{d1}_{mk,32,m} = 2^{-1}\boldsymbol{b}_{km}\{\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, and other elements equal to 0, $\boldsymbol{\tau}^{d1}_{mk,24} = (\boldsymbol{\tau}^{d1\top}_{mk,24,1}, \dots, \boldsymbol{\tau}^{d1\top}_{mk,24,M})^\top = \boldsymbol{\tau}^{d1\top}_{mk,42}$ with $\boldsymbol{\tau}^{d1}_{mk,24,m} = 2^{-1}\boldsymbol{b}_{km}\{\text{vec}(\sum_{i\in\mathcal{I}_K}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top)+ N_k^{-1}\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, and other elements equal to 0, and $\boldsymbol{\tau}^{d1}_{mk,34} = \boldsymbol{\tau}^{d1\top}_{mk,43} = 4^{-1}\mathbf{D}^\top\{\text{vec}($

$\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\mathrm{vec}(\sum_{i\in\mathcal{I}_K}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top)+N_k^{-1}\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$. Combing the results of (A.5), (A.6), and (A.7), we can obtain

$$
\begin{aligned}
&d^2\mathcal{L}_1(\boldsymbol{\theta})+d^2\mathcal{L}_2(\boldsymbol{\theta})+d^2\mathcal{L}_3(\boldsymbol{\theta})\\
=&\Big\{d\mathrm{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top\Big[2^{-1}\sum_{k=1}^K\mathbf{D}^\top\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}\Big]d\mathrm{vech}\big(\boldsymbol{\Omega}\big)\\
&+\Big\{d\mathrm{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top\Big[2^{-1}\sum_{k=1}^K\mathbf{D}^\top\Big\{\big(N_k-1\big)\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\big)\\
&+N_k^{-2}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}\Big]d\mathrm{vech}\big(\boldsymbol{\Sigma}\big)\\
&+\Big\{d\mathrm{vech}\big(\boldsymbol{\Sigma}\big)\Big\}^\top\Big[2^{-1}\sum_{k=1}^K\mathbf{D}^\top\Big\{N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}\Big]d\mathrm{vech}\big(\boldsymbol{\Omega}\big)\\
&+\Big\{d\mathrm{vech}\big(\boldsymbol{\Omega}\big)\Big\}^\top\Big[2^{-1}\sum_{k=1}^K\mathbf{D}^\top\Big\{N_k^{-1}\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\mathbf{D}\Big]d\mathrm{vech}\big(\boldsymbol{\Sigma}\big).
\end{aligned}
$$

According to (A.8), we can compute the $d^2\mathcal{L}_4(\boldsymbol{\theta})$. Therefore, $d^2\mathcal{L}(\boldsymbol{\theta})$ can be proved as

$$
d^2\mathcal{L}\big(\boldsymbol{\theta}\big)=\begin{bmatrix}d\boldsymbol{\alpha}\\ d\boldsymbol{\nu}\\ d\mathrm{vech}(\boldsymbol{\Omega})\\ d\mathrm{vech}(\boldsymbol{\Sigma})\end{bmatrix}^\top\begin{bmatrix}\widehat{\boldsymbol{\Pi}}_{11}&\widehat{\boldsymbol{\Pi}}_{12}&\widehat{\boldsymbol{\Pi}}_{13}&\widehat{\boldsymbol{\Pi}}_{14}\\ \widehat{\boldsymbol{\Pi}}_{21}&\widehat{\boldsymbol{\Pi}}_{22}&\widehat{\boldsymbol{\Pi}}_{23}&\widehat{\boldsymbol{\Pi}}_{24}\\ \widehat{\boldsymbol{\Pi}}_{31}&\widehat{\boldsymbol{\Pi}}_{32}&\widehat{\boldsymbol{\Pi}}_{33}&\widehat{\boldsymbol{\Pi}}_{34}\\ \widehat{\boldsymbol{\Pi}}_{41}&\widehat{\boldsymbol{\Pi}}_{42}&\widehat{\boldsymbol{\Pi}}_{43}&\widehat{\boldsymbol{\Pi}}_{44}\end{bmatrix}\begin{bmatrix}d\boldsymbol{\alpha}\\ d\boldsymbol{\nu}\\ d\mathrm{vech}(\boldsymbol{\Omega})\\ d\mathrm{vech}(\boldsymbol{\Sigma})\end{bmatrix},
$$

where $\widehat{\boldsymbol{\Pi}}_{11}=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top\in\mathbb{R}^{(M-1)\times(M-1)}$, $\widehat{\boldsymbol{\Pi}}_{22}=(\widehat{\boldsymbol{\Pi}}_{22,m_1m_2})=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\nu}^\top\in\mathbb{R}^{Mp\times Mp}$ with $\widehat{\boldsymbol{\Pi}}_{22,m_1m_2}\in\mathbb{R}^{p\times p}$ for $1\le m_1,m_2\le M$, $\widehat{\boldsymbol{\Pi}}_{33}=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\mathrm{vech}(\boldsymbol{\Omega})\partial\mathrm{vech}(\boldsymbol{\Omega})^\top$ $\in\mathbb{R}^{\{p(p+1)/2\}\times\{p(p+1)/2\}}$, and $\widehat{\boldsymbol{\Pi}}_{44}=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\mathrm{vech}(\boldsymbol{\Sigma})\partial\mathrm{vech}(\boldsymbol{\Sigma})^\top\in\mathbb{R}^{\{p(p+1)/2\}\times\{p(p+1)/2\}}$. The off-diagonal block matrices are given by $\widehat{\boldsymbol{\Pi}}_{12}=\widehat{\boldsymbol{\Pi}}_{21}^\top=(\widehat{\boldsymbol{\Pi}}_{12,1},\ldots,\widehat{\boldsymbol{\Pi}}_{12,M})=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\nu}^\top\in\mathbb{R}^{(M-1)\times Mp}$ with $\widehat{\boldsymbol{\Pi}}_{12,m}\in\mathbb{R}^{(M-1)\times p}$ and $1\le m\le M$, $\widehat{\boldsymbol{\Pi}}_{13}=\widehat{\boldsymbol{\Pi}}_{31}^\top=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\mathrm{vech}(\boldsymbol{\Omega})^\top$, $\widehat{\boldsymbol{\Pi}}_{14}=\widehat{\boldsymbol{\Pi}}_{41}^\top=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}\partial\mathrm{vech}(\boldsymbol{\Sigma})^\top\in\mathbb{R}^{(M-1)\times\{p(p+1)/2\}}$, and $\widehat{\boldsymbol{\Pi}}_{23}=\widehat{\boldsymbol{\Pi}}_{32}^\top=(\widehat{\boldsymbol{\Pi}}_{23,1}^\top,\ldots,\widehat{\boldsymbol{\Pi}}_{23,M}^\top)^\top=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\nu}\partial\mathrm{vech}(\boldsymbol{\Omega})^\top\in\mathbb{R}^{Mp\times\{p(p+1)/2\}}$ with $\widehat{\boldsymbol{\Pi}}_{23,m}\in\mathbb{R}^{p\times\{p(p+1)/2\}}$ and $1\le m\le M$. Additionally, $\widehat{\boldsymbol{\Pi}}_{24}=(\widehat{\boldsymbol{\Pi}}_{24,1}^\top,\ldots,\widehat{\boldsymbol{\Pi}}_{24,M}^\top)^\top=\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\nu}\partial\mathrm{vech}(\boldsymbol{\Sigma})^\top=\widehat{\boldsymbol{\Pi}}_{42}^\top\in\mathbb{R}^{Mp\times\{p(p+1)/2\}}$ with $\widehat{\boldsymbol{\Pi}}_{42,m}\in\mathbb{R}^{p\times\{p(p+1)/2\}}$, and $\widehat{\boldsymbol{\Pi}}_{34}=$

$\widehat{\mathbf{\Pi}}_{43}^\top = \partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\mathrm{vech}(\mathbf{\Omega})\partial\mathrm{vech}(\mathbf{\Sigma})^\top \in \mathbb{R}^{\{p(p+1)/2\}\times\{p(p+1)/2\}}$. The analytical form can be given in the following.

$$\widehat{\mathbf{\Pi}}_{11} = -\sum_{k=1}^{K}\Big(\sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\widetilde{\boldsymbol{\alpha}}_{m'}\Big)\Big(\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\widetilde{\boldsymbol{\alpha}}_{m}\Big)^\top,$$

$$\widehat{\mathbf{\Pi}}_{22,mm} = \sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\big(1-\pi_{(\mathcal{I}_k)}^{m}\big)\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top - \sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1},$$

$$\widehat{\mathbf{\Pi}}_{22,m_1m_2} = -\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m_1}\pi_{(\mathcal{I}_k)}^{m_2}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^\top \quad \text{for} \quad m_1\neq m_2,$$

$$\widehat{\mathbf{\Pi}}_{33} = 2^{-1}\sum_{k=1}^{K}\mathbf{D}^\top\Big\{\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\Big\}\mathbf{D} + 4^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}$$
$$\mathbf{D}^\top\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}^\top\mathbf{D}$$
$$-\sum_{k=1}^{K}\sum_{m=1}^{M}\mathbf{D}^\top\pi_{(\mathcal{I}_k)}^{m}\Big\{\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\otimes\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\Big\}\mathbf{D},$$

$$\widehat{\mathbf{\Pi}}_{44} = 2^{-1}\sum_{k=1}^{K}\mathbf{D}^\top\Big\{(N_k-1)\big(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\big) + N_k^{-2}\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\otimes\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\Big\}\mathbf{D}$$
$$+4^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\mathbf{D}^\top\Bigg[\Big\{N_k^{-1}\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}$$
$$\times\Big\{\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}^\top\Bigg]\mathbf{D}$$
$$-\sum_{k=1}^{K}\sum_{m=1}^{M}\mathbf{D}^\top\pi_{(\mathcal{I}_k)}^{m}\Big\{N_k^{-2}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\otimes\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1} + \Big(\sum_{i\in\mathcal{I}_k}\boldsymbol{c}_{ik}\boldsymbol{c}_{ik}^\top\Big)\otimes\mathbf{\Sigma}^{-1}\Big\}\mathbf{D},$$

$$\widehat{\mathbf{\Pi}}_{12,m} = \sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\Big(\widetilde{\boldsymbol{\alpha}}_m - \sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\widetilde{\boldsymbol{\alpha}}_{m'}\Big)\boldsymbol{b}_{km}^\top,$$

$$\widehat{\mathbf{\Pi}}_{13} = 2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\widetilde{\boldsymbol{\alpha}}_m\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top\mathbf{D},$$

$$\widehat{\mathbf{\Pi}}_{14} = 2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m}\widetilde{\boldsymbol{\alpha}}_m\left\{\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\right\}^\top\mathbf{D},$$

$$\widehat{\mathbf{\Pi}}_{23,m} = -\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\Big\{\boldsymbol{b}_{km}^\top\otimes(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\Big\}\mathbf{D}$$
$$+2^{-1}\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m}\boldsymbol{b}_{km}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M}\pi_{(\mathcal{I}_k)}^{m'}\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top\mathbf{D},$$

$$
\begin{aligned}
\widehat{\mathbf{\Pi}}_{24,m} = & -\sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m} \Big\{ \boldsymbol{b}_{km}^{\top} \otimes N_k^{-1} (N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1} \Big\} \mathbf{D} \\
& + 2^{-1} \sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m} \boldsymbol{b}_{km} \Bigg\{ \mathrm{vec}\Big(N_k^{-1} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big) - \sum_{m'=1}^{M} \pi_{(\mathcal{I}_k)}^{m'} \mathrm{vec}\Big(N_k^{-1} \boldsymbol{b}_{km'} \boldsymbol{b}_{km'}^{\top} \Big) \Bigg\}^{\top} \mathbf{D}, \\
\widehat{\mathbf{\Pi}}_{34} = & \, 2^{-1} \sum_{k=1}^{K} \mathbf{D}^{\top} \Big\{ N_k^{-1} \big(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega} \big)^{-1} \otimes \big(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega} \big)^{-1} \Big\} \mathbf{D} \\
& + 4^{-1} \sum_{k=1}^{K} \sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \mathbf{D}^{\top} \Big\{ \mathrm{vec}\Big(\boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big) \Big\} \Bigg[\mathrm{vec}\Big(N_k^{-1} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big) - \sum_{m'=1}^{M} \pi_{(\mathcal{I}_k)}^{m'} \Big\{ \\
& \mathrm{vec}\Big(N_k^{-1} \boldsymbol{b}_{km'} \boldsymbol{b}_{km'}^{\top} \Big) \Big\} \Bigg]^{\top} \mathbf{D} - \sum_{k=1}^{K} \sum_{m=1}^{M} \pi_{(\mathcal{I}_k)}^{m} \mathbf{D}^{\top} \Big\{ \Big(N_k^{-1} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big) \otimes \big(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega} \big)^{-1} \Big\} \mathbf{D}.
\end{aligned}
$$

STEP 2. We consider computing $\mathbf{\Pi}^{N,K}$ by evaluating every block matrices $E(\widehat{\mathbf{\Pi}}_{j_1 j_2})$ separately, for $1 \leq j_1, j_2 \leq 4$. To begin with, the expectation of $\widehat{\mathbf{\Pi}}_{11}$ can be given as

$$
E(\widehat{\mathbf{\Pi}}_{11}) = -\sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{m'=1}^{M} E\Big(\pi_{(\mathcal{I}_k)}^{m} \pi_{(\mathcal{I}_k)}^{m'} \Big) \widetilde{\boldsymbol{\alpha}}_m \widetilde{\boldsymbol{\alpha}}_{m'}^{\top}, \tag{A.9}
$$

Then, by decomposing $\widehat{\mathbf{\Pi}}_{22,mm} = \widehat{\mathbf{\Pi}}_{22,mm}^{(1)} - \widehat{\mathbf{\Pi}}_{22,mm}^{(2)}$, where $\widehat{\mathbf{\Pi}}_{22,mm}^{(1)} = \sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m} (1 - \pi_{(\mathcal{I}_k)}^{m}) \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top}$ and $\widehat{\mathbf{\Pi}}_{22,mm}^{(2)} = \sum_{k=1}^{K} \pi_{(\mathcal{I}_k)}^{m} (N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}$, we can compute its expectation individually. We first study the second term $E(\widehat{\mathbf{\Pi}}_{22,mm}^{(2)}) = \sum_{k=1}^{K} E(\pi_{(\mathcal{I}_k)}^{m})(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}$. Since $E(\pi_{(\mathcal{I}_k)}^{m}) = \alpha_m$, we then have $E(\widehat{\mathbf{\Pi}}_{22,mm}^{(2)}) = \sum_{k=1}^{K} \alpha_m (N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}$. Subsequently, we need to calculate $E(\widehat{\mathbf{\Pi}}_{22,mm}^{(1)}) = \sum_{k=1}^{K} E(\pi_{(\mathcal{I}_k)}^{m} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top}) - \sum_{k=1}^{K} E\{(\pi_{(\mathcal{I}_k)}^{m})^2 \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top}\}$, where $E(\pi_{(\mathcal{I}_k)}^{m} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top})$ can be proved as

$$
\begin{aligned}
E\Big(\pi_{(\mathcal{I}_k)}^{m} \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big) = & E\Big[E\Big\{ \mathbf{1}\big(Z_k = m \big) \boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big| \mathbf{X}_{(\mathcal{I}_k)}, \boldsymbol{Y}_{(\mathcal{I}_k)} \Big\} \Big] \\
= & \alpha_m E\Big(\boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} \Big| Z_k = m \Big),
\end{aligned} \tag{A.10}
$$

$E(\boldsymbol{b}_{km} \boldsymbol{b}_{km}^{\top} | Z_k = m) = (N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1} E\{(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)^{\top} | Z_k = m\} (N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}$, and $E\{(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m)^{\top} | Z_k = m\}$ is given as

$$
E\Big\{ \big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m \big) \big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m \big)^{\top} \Big| Z_k = m \Big\}
$$

$$=E\left[\left\{\left(\overline{\boldsymbol{X}}_k-\boldsymbol{\mu}_{km}+\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m\right)\left(\overline{\boldsymbol{X}}_k-\boldsymbol{\mu}_{km}+\boldsymbol{\mu}_{km}-\boldsymbol{\nu}_m\right)^\top\middle|Z_k=m,\boldsymbol{\mu}_{km}\right\}\middle|Z_k=m\right]$$
$$=N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}. \tag{A.11}$$

Therefore, $E(\widehat{\boldsymbol{\Pi}}_{22,mm}^{(1)})=\sum_{k=1}^K\alpha_m(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}-\sum_{k=1}^K E\{(\pi_{(\mathcal{I}_k)}^m)^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\}$. By combing the results above, we can obtain that

$$E(\widehat{\boldsymbol{\Pi}}_{22,mm})=-\sum_{k=1}^K E\Big\{\big(\pi_{(\mathcal{I}_k)}^m\big)^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big\}. \tag{A.12}$$

We also can prove that the expectation of $\widehat{\boldsymbol{\Pi}}_{22,m_1m_2}$ with $1\le m_1\ne m_2\le M$ is

$$E(\widehat{\boldsymbol{\Pi}}_{22,m_1m_2})=-\sum_{k=1}^K E\Big(\pi_{(\mathcal{I}_k)}^{m_1}\pi_{(\mathcal{I}_k)}^{m_2}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^\top\Big). \tag{A.13}$$

Next, we consider computing $E(\widehat{\boldsymbol{\Pi}}_{33})=E(\widehat{\boldsymbol{\Pi}}_{33}^{(1)})+E(\widehat{\boldsymbol{\Pi}}_{33}^{(2)})-E(\widehat{\boldsymbol{\Pi}}_{33}^{(3)})$, where $E(\widehat{\boldsymbol{\Pi}}_{33}^{(1)})=\widehat{\boldsymbol{\Pi}}_{33}^{(1)}=2^{-1}\sum_{k=1}^K\mathbf{D}^\top\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\otimes(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\mathbf{D}$, $\widehat{\boldsymbol{\Pi}}_{33}^{(2)}=4^{-1}\sum_{k=1}^K\sum_{m=1}^M\pi_{(\mathcal{I}_k)}^m\mathbf{D}^\top$ $\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)-\sum_{m'=1}^M\pi_{(\mathcal{I}_k)}^{m'}\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top)\}\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top\mathbf{D}$, and $\widehat{\boldsymbol{\Pi}}_{33}^{(3)}=\sum_{k=1}^K\sum_{m=1}^M$ $\mathbf{D}^\top\pi_{(\mathcal{I}_k)}^m\{(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\otimes(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\mathbf{D}$. Similarly as (A.10), (A.11), and (A.12), we have $E(\widehat{\boldsymbol{\Pi}}_{33}^{(3)})=\sum_{k=1}^K\mathbf{D}^\top\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\otimes(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\mathbf{D}$. Since it can be verified that $E[\pi_{(\mathcal{I}_k)}^m\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}^\top]=2\alpha_m\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\otimes(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}+\alpha_m\mathrm{vec}\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}\mathrm{vec}^\top\{(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\}$, we have $E(\widehat{\boldsymbol{\Pi}}_{33}^{(2)})=2^{-1}E(\widehat{\boldsymbol{\Pi}}_{33}^{(3)})-4^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{m'=1}^M\mathbf{D}^\top E[\pi_{(\mathcal{I}_k)}^m\pi_{(\mathcal{I}_k)}^{m'}\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top)\}^\top]\mathbf{D}$. Thus, the expectation $E(\widehat{\boldsymbol{\Pi}}_{33})$ can be proved as

$$\begin{aligned}E(\widehat{\boldsymbol{\Pi}}_{33})=&4^{-1}\sum_{k=1}^K\mathbf{D}^\top\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\Big[\mathrm{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\Big]^\top\mathbf{D}\\&-4^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{m'=1}^M\mathbf{D}^\top E\left[\pi_{(\mathcal{I}_k)}^m\pi_{(\mathcal{I}_k)}^{m'}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top\right]\mathbf{D}.\end{aligned} \tag{A.14}$$

Then, we consider studying $E(\widehat{\boldsymbol{\Pi}}_{44})$. With similar and complex calculation as showed in (A.10), (A.11), (A.12) and (A.14), we can obtain that

$$
\begin{aligned}
E(\widehat{\mathbf{\Pi}}_{44}) = & -2^{-1}\sum_{k=1}^{K}\mathbf{D}^{\top}\Big\{(N_k-1)\big(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\big)\Big\}\mathbf{D}+4^{-1}\sum_{k=1}^{K}\mathbf{D}^{\top}\Big[N_k^{-2} \\
& \times \mathrm{vec}\Big\{\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\Big\}\mathrm{vec}\Big\{\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\Big\}^{\top}\Big]\mathbf{D} \\
& -4^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbf{D}^{\top}E\Big[\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\Big\{N_k^{-1}\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\Big)\Big\} \\
& \times\Big\{N_k^{-1}\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top}\Big)\Big\}^{\top}\Big]\mathbf{D}.
\end{aligned}
\tag{A.15}
$$

Next, we continuing computing the expectations of these off-diagonal block matrices as follows. We first calculate $E(\widehat{\mathbf{\Pi}}_{12}) = \Big(E(\widehat{\mathbf{\Pi}}_{12,1}),\ldots,E(\widehat{\mathbf{\Pi}}_{12,M})\Big)$, where $E(\widehat{\mathbf{\Pi}}_{12,m}) = E(\widehat{\mathbf{\Pi}}^{(1)}_{12,m})-E(\widehat{\mathbf{\Pi}}^{(2)}_{12,m})$, $E(\widehat{\mathbf{\Pi}}^{(1)}_{12,m}) = \sum_{k=1}^{K}\widetilde{\boldsymbol{\alpha}}_m E(\pi^{m}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^{\top})$, and $E(\widehat{\mathbf{\Pi}}^{(2)}_{12,m}) = \sum_{k=1}^{K}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_{m'}E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^{\top})$. It can be proved that $E(\pi^{m}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^{\top}) = \mathbf{0}$, we have $E(\widehat{\mathbf{\Pi}}^{(1)}_{12,m}) = \mathbf{0}$. Therefore, it can be verfied that

$$
E(\widehat{\mathbf{\Pi}}_{12,m}) = -\sum_{k=1}^{K}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_{m'}E\Big(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^{\top}\Big). \tag{A.16}
$$

To evaluate $E(\widehat{\mathbf{\Pi}}_{13}) = E(\widehat{\mathbf{\Pi}}^{(1)}_{13})-E(\widehat{\mathbf{\Pi}}^{(2)}_{13})$, we compute $E(\widehat{\mathbf{\Pi}}^{(1)}_{13}) = 2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\widetilde{\boldsymbol{\alpha}}_m$ $E[\pi^{m}_{(\mathcal{I}_k)}\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top})\}^{\top}]\mathbf{D}$, and $E(\widehat{\mathbf{\Pi}}^{(2)}_{13}) = 2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}E[\pi^{m}_{(\mathcal{I}_k)}\widetilde{\boldsymbol{\alpha}}_m\{\sum_{m'=1}^{M}\pi^{m'}_{(\mathcal{I}_k)}\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top})\}^{\top}]\mathbf{D}$, separaterly. Since $\sum_{m=1}^{M}\alpha_m\widetilde{\boldsymbol{\alpha}}_m = \mathbf{0}$, we then have $E(\widehat{\mathbf{\Pi}}^{(1)}_{13}) = 2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\alpha_m\widetilde{\boldsymbol{\alpha}}_m\{(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\otimes(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\}\mathbf{D} = \mathbf{0}$. Then, it can be verified that

$$
E(\widehat{\mathbf{\Pi}}_{13}) = -2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_m E\Big[\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top}\Big)\Big\}^{\top}\Big]\mathbf{D}. \tag{A.17}
$$

By similar decomposing and calculation, we can obtain that

$$
E(\widehat{\mathbf{\Pi}}_{14}) = -2^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_m E\Big[\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\Big\{\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top}\Big)\Big\}^{\top}\Big]\mathbf{D}. \tag{A.18}
$$

Next, we consider computing $E(\widehat{\mathbf{\Pi}}_{23}) = (E(\widehat{\mathbf{\Pi}}_{23,1})^{\top},\ldots,E(\widehat{\mathbf{\Pi}}_{23,M})^{\top})^{\top}$, where $E(\widehat{\mathbf{\Pi}}_{23,m}) = -E(\widehat{\mathbf{\Pi}}^{(1)}_{23,m})+E(\widehat{\mathbf{\Pi}}^{(2)}_{23,m})-E(\widehat{\mathbf{\Pi}}^{(3)}_{23,m})$, where $E(\widehat{\mathbf{\Pi}}^{(1)}_{23,m}) = \sum_{k=1}^{K}\{E(\pi^{m}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^{\top})\otimes(N_k^{-1}\mathbf{\Sigma}+$

$\boldsymbol{\Omega})^{-1}\}\mathbf{D} = \mathbf{0}$, $E(\widehat{\boldsymbol{\Pi}}_{23,m}^{(2)}) = 2^{-1}\sum_{k=1}^{K} E[\pi_{(\mathcal{I}_k)}^{m}\boldsymbol{b}_{km}\{\text{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top})\}^{\top}]\mathbf{D} = \mathbf{0}$, and $E(\widehat{\boldsymbol{\Pi}}_{23,m}^{(3)}) = 2^{-1}\sum_{k=1}^{K}\sum_{m'=1}^{M} E[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\boldsymbol{b}_{km}\{\text{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top})\}^{\top}]\mathbf{D}$. Therefore, we can obtain that

$$E(\widehat{\boldsymbol{\Pi}}_{23,m}) = -2^{-1}\sum_{k=1}^{K}\sum_{m'=1}^{M} E\left[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\boldsymbol{b}_{km}\left\{\text{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top}\Big)\right\}^{\top}\right]\mathbf{D}. \tag{A.19}$$

By similar decomposing and calculation, it can verified that $E(\widehat{\boldsymbol{\Pi}}_{24}) = (E(\widehat{\boldsymbol{\Pi}}_{24,1})^{\top}, \ldots, E(\widehat{\boldsymbol{\Pi}}_{24,M})^{\top})^{\top}$, where

$$E(\widehat{\boldsymbol{\Pi}}_{24,m}) = -2^{-1}\sum_{k=1}^{K}\sum_{m'=1}^{M} E\left[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\boldsymbol{b}_{km}\left\{\text{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top}\Big)\right\}^{\top}\right]\mathbf{D}. \tag{A.20}$$

Following similar calculation (A.10), (A.11), (A.12) and (A.14), we have

$$E(\widehat{\boldsymbol{\Pi}}_{34}) = 4^{-1}\sum_{k=1}^{K}\mathbf{D}^{\top}\left[N_k^{-1}\text{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}\text{vec}\Big\{\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\Big\}^{\top}\right]\mathbf{D}-$$
$$4^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbf{D}^{\top}E\left[\pi_{(\mathcal{I}_k)}^{m}\pi_{(\mathcal{I}_k)}^{m'}\left\{\text{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\Big)\right\}\left\{N_k^{-1}\text{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^{\top}\Big)\right\}^{\top}\right]\mathbf{D}. \tag{A.21}$$

The expectation of $\ddot{\mathcal{L}}(\boldsymbol{\theta})$ have been gained. Comparing the above results with those in Appendix A.3, we can verify that $E\{\ddot{\mathcal{L}}(\boldsymbol{\theta})\} = -\text{cov}\{\dot{\mathcal{L}}(\boldsymbol{\theta})\}$.

APPENDIX B

Appendix B.1: Proof of Theorem 1

Subsequently, we study the asymptotic form of $E(\widehat{\boldsymbol{\Pi}}^{\mathbf{S}}) = \mathbf{S}E\{\ddot{\mathcal{L}}(\boldsymbol{\theta})\}\mathbf{S} = E(\widehat{\boldsymbol{\Pi}}_{j_1j_2}^{\mathbf{S}})$, for $1 \le j_1, j_2 \le 4$ under Conditions (C1)-(C5). Next, we will separately demonstrate the divergence rates of each component as specified in Theorem 1. Particularly, the matrix $\boldsymbol{\Gamma}_{44}$ diverges at the rate of N, and the other matrice diverge at the rate of K. Lastly, we compute the asymptotic form for $E(\widehat{\boldsymbol{\Pi}}^{\mathbf{S}})$.

Step 1. We shall prove the divergence rate of $E(\widehat{\boldsymbol{\Pi}}_{44}^{\mathbf{S}})$ as an example. According to (A.15), we have $E(\widehat{\boldsymbol{\Pi}}_{44}^{\mathbf{S}}) = E(\widehat{\boldsymbol{\Pi}}_{44,1}^{\mathbf{S}}) + E(\widehat{\boldsymbol{\Pi}}_{44,2}^{\mathbf{S}}) + E(\widehat{\boldsymbol{\Pi}}_{44,3}^{\mathbf{S}})$, where $E(\widehat{\boldsymbol{\Pi}}_{44,1}^{\mathbf{S}}) =$

$-2^{-1}N^{-1}\sum_{k=1}^K \mathbf{D}^\top\big\{(N_k-1)(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\big\}\mathbf{D}$, $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,2}) = 4^{-1}N^{-1}\sum_{k=1}^K \mathbf{D}^\top\big[N_k^{-2}\mathrm{vec}\{(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\}\mathrm{vec}\{(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\}^\top\big]\mathbf{D}$, and $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,3}) = -4^{-1}N^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{m'=1}^M \mathbf{D}^\top E\big[\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\{N_k^{-1}\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{N_k^{-1}\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top)\}^\top\big]\mathbf{D}$. Then, it can be proved that $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,1}) = (2N)^{-1}(N-K)\mathbf{D}^\top\big(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\big)\mathbf{D}$. Therefore, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,1})\|_F = O(1)$, where $\|A\|_F$ denotes the Frobenius norm of the matrix A. Since $\|N_k^{-1}\mathbf{\Sigma}\|_F = o(1)$, $\|\mathbf{\Omega}^{-1}\|_F = O(1)$ and $K = o(N)$ as $N\to\infty$, $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,2}) = o(1)$. By a similar argument, it can be shown that $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,3})\|_F = o(1)$. With the triangle inequality, we have $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44})\|_F \le \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,1})\|_F + \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,2})\|_F + \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,3})\|_F = O(1)$.

Step 2. Here, we shall prove that other block matrice diverge at the rate of K. To begin with, we compute the other on-diagonal matrice of $\widehat{\mathbf{\Pi}}^{\mathbf{S}}$. As demonstrated in (A.9), we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11}) = -K^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{m'=1}^M E\big(\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\big)\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_{m'}^\top$. Therefore, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11})\|_F \le \|K^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{m'=1}^M\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_{m'}^\top\|_F = O(1)$. With (A.13), we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2}) = -K^{-1}\sum_{k=1}^K E\big(\pi^{m_1}_{(\mathcal{I}_k)}\pi^{m_2}_{(\mathcal{I}_k)}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^\top\big)$. Then, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2})\|_F \le \|K^{-1}\sum_{k=1}^K E(\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^\top)\|_F = O(1)$. Similarly, we can obtain that $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) = E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,1}) + E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,2})$, where $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,1} = 4K^{-1}\sum_{k=1}^K\mathbf{D}^\top\mathrm{vec}\{(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\}\big[\mathrm{vec}\{(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}\}\big]^\top\mathbf{D}$, $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,2} = -4K^{-1}\sum_{k=1}^K\sum_{m=1}^M\sum_{m'=1}^M\mathbf{D}^\top E\big[\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\{\mathrm{vec}(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top)\}\{\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top)\}^\top\big]\mathbf{D}$. Since $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,1})\|_F = O(1)$ and $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,2})\|_F = O(1)$, we then have $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33})\|_F \le \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,1})\|_F + \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33,2})\|_F = O(1)$ by applying the triangle inequality. Next, we calculate the off-diagonal block matrice of $\widehat{\mathbf{\Pi}}^{\mathbf{S}}$. By similarly arguments and the results (A.16), (A.17), (A.18), (A.19), (A.20), and (A.21) mentioned above, we can prove that $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12})\|_F = \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{21})\|_F = O(1)$, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{13})\|_F = \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{31})\|_F = O(1)$, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{14})\|_F = \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{41})\|_F = O(1)$, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{23})\|_F = \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{32})\|_F = O(1)$, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{24})\|_F = \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{42})\|_F = O(1)$, $\|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{34})\|_F = \|E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{43})\|_F = O(1)$.

Step 3. Subsequently, we calculate the asymptotic form of $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}})$. We start with computing the on-diagonal matrice of $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}})$. As mentioned in the Step 1, $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44}) = E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,1} + \widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,2} + \widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,3})$, where $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,1}) \to -2^{-1}\mathbf{D}^\top\big(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\big)\mathbf{D}$, $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,2}) = o(1)$, and $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44,3}) = o(1)$. Therefore, we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44}) \to -2^{-1}\mathbf{D}^\top\big(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\big)\mathbf{D} = -\mathbf{\Pi}_{44}$.

Next, we calculate the asymptotic forms of the other on-diagonal matrices. For

$E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11})$, as mentioned in A.9, we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11}) = -K^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M} E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)})$ $\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_{m'}^{\top}$, with the specific form of $\pi^{m}_{(\mathcal{I}_k)}$ as

$$\pi^{m}_{(\mathcal{I}_k)} = \Big[\alpha_m \exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)^{\top}\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m\big)\Big\}\Big]$$
$$\Big[\sum_{m'=1}^{M}\alpha_{m'} \exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_{m'}\big)^{\top}\big(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_{m'}\big)\Big\}\Big]^{-1}.$$

Therefore, we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11}) = -K^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M} E\left[E\left(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}|\mathbb{B}\right)\right]\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_{m'}^{\top}$. Since samples within the same class are independent and identically distributed with both mean and variance existing, we can apply Chebyshev's Law of Large Numbers to conclude that $\overline{\boldsymbol{X}}_k$ converges in probability to the corresponding class mean $\boldsymbol{\mu}_{kZ_k}$ with a known $\mathbb{B}$ as $N_K \to \infty$. Furthermore, since $\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)} \le 1$ is a bounded and continuous function of $\overline{\boldsymbol{X}}_k$, by applying the Portmanteau Lemma, we can conclude that $E\left(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}|\mathbb{B}\right)$ converges uniformly to $E\left(\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}|\mathbb{B}\right)$, where

$$\pi^{m*}_{(\mathcal{I}_k)} = \Big[\sum_{m'=1}^{M}\alpha_{m'} \exp\Big\{-2^{-1}\big(\boldsymbol{\mu}_{kZ_k} - \boldsymbol{\nu}_{m'}\big)^{\top}\mathbf{\Omega}^{-1}\big(\boldsymbol{\mu}_{kZ_k} - \boldsymbol{\nu}'_{m}\big)\Big\}\Big]^{-1}\alpha_m \exp\Big\{-2^{-1}$$
$$\big(\boldsymbol{\mu}_{kZ_k} - \boldsymbol{\nu}_{m}\big)^{\top}\mathbf{\Omega}^{-1}\big(\boldsymbol{\mu}_{kZ_k} - \boldsymbol{\nu}_{m}\big)\Big\}.$$

Therefore, we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11}) \to -\sum_{m=1}^{M}\sum_{m'=1}^{M} E\left(\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\right)\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_{m'}^{\top} = -\mathbf{\Pi}_{11}$, where $\to$ denotes uniform convergence.

For $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22})$, the asymptotic forms of $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,mm}) = -K^{-1}\sum_{k=1}^{K} E\{(\pi^{m}_{(\mathcal{I}_k)})^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\}$ and $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2}) = -K^{-1}\sum_{k=1}^{K} E(\pi^{m_1}_{(\mathcal{I}_k)}\pi^{m_2}_{(\mathcal{I}_k)}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^{\top})$ can be obtained. To illustrate the methodology for deriving these asymptotic forms, we will use the more general part $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2})$ as an example. Recall that $\boldsymbol{b}_{km} = (N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega})^{-1}(\overline{\boldsymbol{X}}_k - \boldsymbol{\nu}_m) \in \mathbb{R}^p$. Notably, $\pi^{m_1}_{(\mathcal{I}_k)}\pi^{m_2}_{(\mathcal{I}_k)}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^{\top}$ constitutes a bounded continuous function of $\overline{\boldsymbol{X}}_k$. This property permits application of the Portmanteau Lemma, yielding the conditional expectation convergence $E(\pi^{m_1}_{(\mathcal{I}_k)}\pi^{m_2}_{(\mathcal{I}_k)}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^{\top}|\mathbb{B}) \to E(\pi^{m_1*}_{(\mathcal{I}_k)}\pi^{m_2*}_{(\mathcal{I}_k)}\boldsymbol{b}^{*}_{km_1}\boldsymbol{b}^{*\top}_{km_2}|\mathbb{B})$, where $\boldsymbol{b}^{*}_{km} = \mathbf{\Omega}^{-1}(\boldsymbol{\mu}_{kZ_k} - \boldsymbol{\nu}_m)$. By utilizing the Law of Total Expectation, we have $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2}) = -K^{-1}\sum_{k=1}^{K} E\left[E(\pi^{m_1}_{(\mathcal{I}_k)}\pi^{m_2}_{(\mathcal{I}_k)}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^{\top}|\mathbb{B})\right] \to -E(\pi^{m_1*}_{(\mathcal{I}_k)}\pi^{m_2*}_{(\mathcal{I}_k)}\boldsymbol{b}^{*}_{km_1}\boldsymbol{b}^{*\top}_{km_2}) = -\mathbf{\Pi}_{22,m_1,m_2}$.

Similarly, it can be verified that $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2}) \to -\mathbf{\Pi}_{22,m,m}$.

Next, we address $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) = E_1(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) + E_2(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33})$, with the specific forms given as follows:

$$\begin{aligned}
E_1(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) =& 4K^{-1}\sum_{k=1}^{K}\mathbf{D}^\top \mathrm{vec}\left\{\left(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\right)^{-1}\right\}\left[\mathrm{vec}\left\{\left(N_k^{-1}\mathbf{\Sigma}+\mathbf{\Omega}\right)^{-1}\right\}\right]^\top\mathbf{D}\\
E_2(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) =& -4K^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbf{D}^\top E\left[\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\left\{\mathrm{vec}\left(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\right)\right\}\left\{\mathrm{vec}\left(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\right)\right\}^\top\right]\mathbf{D}.
\end{aligned}$$

It can be easily verified that $E_1(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) \to 4\mathbf{D}^\top\mathrm{vec}\left(\mathbf{\Omega}^{-1}\right)\left[\mathrm{vec}\left(\mathbf{\Omega}^{-1}\right)\right]^\top\mathbf{D}$ as $N_K \to \infty$. For $E_2(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33})$, we follow a similar procedure as used for $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11})$. Specifically, by reformulating the expectation part into a double expectation form conditioned on set $\mathbb{B}$ and leveraging the Law of Large Numbers together with the Portmanteau Lemma, we can rigorously prove that $E\left[\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\left\{\mathrm{vec}\left(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\right)\right\}\left\{\mathrm{vec}\left(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\right)\right\}^\top\right] \to E\big[\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}$ $\left\{\mathrm{vec}\left(\boldsymbol{b}^*_{km}\boldsymbol{b}^{*\top}_{km}\right)\right\}\left\{\mathrm{vec}\left(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\right)\right\}^\top\big]$, as $N_k \to \infty$. Under the given conditions, the above approach allows us to demonstrate that

$$E_2(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) \to -4^{-1}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbf{D}^\top E\big[\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\left\{\mathrm{vec}\left(\boldsymbol{b}^*_{km}\boldsymbol{b}^{*\top}_{km}\right)\right\}\left\{\mathrm{vec}\left(\boldsymbol{b}^*_{km'}\boldsymbol{b}^{*\top}_{km'}\right)\right\}^\top\big]\mathbf{D}.$$

Therefore, combining the results of $E_1(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33})$ and $E_2(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33})$, we can obtain that $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}) \to -\mathbf{\Pi}_{33}$, where $\mathbf{\Pi}_{33}$ is mentioned in Theorem 1, as $N_k \to \infty$.

Additionally, we consider handling the off-diagonal block matrices. First, we compute $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12}) = \left(E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12,1}),\dots,E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12,M})\right) \in \mathbb{R}^{(M-1)\times MP}$, Specifically, the expectation of $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12,m}$ can be rewritten as

$$\begin{aligned}
E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12,m}) =& -K^{-1}\sum_{k=1}^{K}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_{m'}E\left(\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^\top\right)\\
=& -K^{-1}\sum_{k=1}^{K}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_{m'}E\left[E\left(\pi^m_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}^\top\Big|\mathbb{B}\right)\right].
\end{aligned}$$

With the Law of Large Numbers, it can be verified that $\overline{\boldsymbol{X}}_k \to \boldsymbol{\mu}_{kZ_k}$ as $N_k \to \infty$.

Applying the Portmanteau Lemma, we can show that

$$E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{12,m}) \to -\sum_{m'=1}^{M} \widetilde{\boldsymbol{\alpha}}_{m'} E\Big(\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\boldsymbol{b}^{*\top}_{km}\Big) = -\mathbf{\Pi}_{12,m}.$$

For $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{13})$, by applying the same methodology, we can obtain that $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{13}) = -(2k)^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_m E\Big\{E\big[\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\{\mathrm{vec}(\boldsymbol{b}_{km'}\boldsymbol{b}^{\top}_{km'})\}^{\top}\big|\mathbb{B}\big]\Big\}\mathbf{D} \to -2^{-1}\sum_{m=1}^{M}\sum_{m'=1}^{M}\widetilde{\boldsymbol{\alpha}}_m E\Big\{E\big[\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\{\mathrm{vec}(\boldsymbol{b}^{*}_{km'}\boldsymbol{b}^{*\top}_{km'})\}^{\top}\big|\mathbb{B}\big]\Big\}\mathbf{D}$, as $N_k \to \infty$. Next, we consider handling $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{23}) = (E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{23,1})^{\top},\dots,E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{23,M})^{\top})^{\top}$. With similar process, as $N_k \to \infty$, it can be verified that

$$E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{23,m}) \to -2^{-1}\sum_{m'=1}^{M} E\Big[\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\boldsymbol{b}^{*}_{km}\left\{\mathrm{vec}(\boldsymbol{b}^{*}_{km'}\boldsymbol{b}^{*\top}_{km'})\right\}^{\top}\Big]\mathbf{D} = -\mathbf{\Pi}_{23,m}.$$

With $N_k \to \infty$, the asymptotic forms of the off-diagonal matrices related to $\mathbf{\Sigma}$ can be easily verified as $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{14}) \to \mathbf{0}$, $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{24}) \to \mathbf{0}$, and $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{34}) \to \mathbf{0}$. At this juncture, we have successfully completed all the necessary proofs for Theorem 1.

Appendix B.2: Proof of Convergence for $\widehat{\mathbf{\Pi}}^{\mathbf{S}}$

In this section, we will demonstrate that $\widehat{\mathbf{\Pi}}^{\mathbf{S}} = \mathbf{S}\ddot{\mathcal{L}}(\boldsymbol{\theta})\,\mathbf{S} \to_p -\mathbf{\Pi}$ under Conditions (C1)-(C5), where the specific form of $\mathbf{\Pi}$ is given in Theorem 1.

Starting with the on-diagonal block matrices, we will proceed by analyzing the large matrix in block form. As mentioned in Appendix A3, we have obtained the specific formula of $\ddot{\mathcal{L}}(\boldsymbol{\theta}) = \widehat{\mathbf{\Pi}}$. Firstly, we handle $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11}$, which is given by $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11} = -K^{-1}\sum_{k=1}^{K}(\sum_{m'=1}^{M}\pi^{m'}_{(\mathcal{I}_k)}\widetilde{\boldsymbol{\alpha}}_{m'})(\sum_{m=1}^{M}\pi^{m}_{(\mathcal{I}_k)}\widetilde{\boldsymbol{\alpha}}_{m})^{\top}$. Due to the independence among $\overline{\boldsymbol{X}}_k$, $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11}$ can be viewed as the average of K independent but not necessarily identically distributed random variables, each with a well-defined mean and finite variance. Therefore, by applying Chebyshev's Law of Large Numbers, we can obtain that $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11} \to^p -K^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M}E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)})\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}^{\top}_{m'}$, as $K \to \infty$. Furthermore, considering that $\overline{\boldsymbol{X}}_k|\mathbb{B} \to \boldsymbol{\mu}_{kZ_k}$ as $N_k \to \infty$, and given that $\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)} \leq 1$ is a bounded and continuous function of $\overline{\boldsymbol{X}}_k$, we can apply the Portmanteau Lemma to conclude that $E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}|\mathbb{B}) \to E(\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)} \mid \mathbb{B})$, as $N_k \to \infty$. Combing the above results, by rewriting $E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)})$ as the double expectation form, we have $E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}) = E\left[E(\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}|\mathbb{B})\right] \to E(\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)})$. Therefore, we have

$$\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{11} \to^{p} -\sum_{m=1}^{M}\sum_{m'=1}^{M} E\left(\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\right)\widetilde{\boldsymbol{\alpha}}_m\widetilde{\boldsymbol{\alpha}}_{m'}^{\top} = -\mathbf{\Pi}_{11}.$$

Next, we proceed to analyze the convergence properties of $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22}$. According to the calculations in Appendix A.3, we further decompose $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22}$ into $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,mm}$ and $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2}$ for a detailed analysis. Specifically, $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,mm} = K^{-1}\sum_{k=1}^{K}\left[\pi^{m}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top} - \pi^{m}_{(\mathcal{I}_k)}(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}\right] - K^{-1}\sum_{k=1}^{K}\left(\pi^{m}_{(\mathcal{I}_k)}\right)^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}$. By applying Chebyshev's Law of Large Numbers, it can be similarly verified that as $K \to \infty$,$\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,mm} \to^{p} K^{-1}\sum_{k=1}^{K} E\Big[\pi^{m}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top} - \pi^{m}_{(\mathcal{I}_k)}(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}\Big] - K^{-1}\sum_{k=1}^{K} E\Big[\left(\pi^{m}_{(\mathcal{I}_k)}\right)^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\Big]$. Recall that $E\big[\pi^{m}_{(\mathcal{I}_k)}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\big] = \alpha_m(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}$, as calculated in Appendix A.3. Therefore, we can obtain that $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,mm} \to^{p} -K^{-1}\sum_{k=1}^{K} E\Big\{E\big[(\pi^{m}_{(\mathcal{I}_k)})^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top}\big|\mathbb{B}\big]\Big\}$. Since $(\pi^{m}_{(\mathcal{I}_k)})^2\boldsymbol{b}_{km}$ is a bounded and continuous function of $\overline{\boldsymbol{X}}_k$, through a similar proof process, we can conclude that $E\big[(\pi^{m}_{(\mathcal{I}_k)})^2\boldsymbol{b}_{km}\boldsymbol{b}_{km}^{\top} \mid \mathbb{B}\big] \to E\big[(\pi^{m*}_{(\mathcal{I}_k)})^2\boldsymbol{b}^{*}_{km}\boldsymbol{b}^{*\top}_{km} \mid \mathbb{B}\big]$, as $N_k \to \infty$, by applying the Portmanteau Lemma. Consequently, we can conclude that

$$\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,mm} \to^{p} -E\left[\left(\pi^{m*}_{(\mathcal{I}_k)}\right)^2\boldsymbol{b}^{*}_{km}\boldsymbol{b}^{*\top}_{km}\right] = -\mathbf{\Pi}_{22,mm}.$$

Similarly, the asymptotic result for $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2} = -K^{-1}\sum_{k=1}^{K}\pi^{m_1}_{(\mathcal{I}_k)}\pi^{m_2}_{(\mathcal{I}_k)}\boldsymbol{b}_{km_1}\boldsymbol{b}_{km_2}^{\top}$ can be derived in the following

$$\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{22,m_1m_2} \to^{p} -E\left(\pi^{m_1*}_{(\mathcal{I}_k)}\pi^{m_2*}_{(\mathcal{I}_k)}\boldsymbol{b}^{*}_{km_1}\boldsymbol{b}^{*\top}_{km_2}\right) = -\mathbf{\Pi}_{22,m_1m_2}.$$

Next, we consider the handling of $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33}$. Based on the specific form of $\widehat{\mathbf{\Pi}}_{33}$ calculated in the Appendix A.3, we observe that all conditions for applying Chebyshev's Law of

Large Numbers are satisfied. Therefore, as $K \to \infty$, we can obtain

$$
\begin{aligned}
\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33} \to^p & \widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,1} + \widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,2} - \widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,3}, \\
\widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,1} = & (2K)^{-1} \sum_{k=1}^{K} \mathbf{D}^\top E \left\{ \left(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\right)^{-1} \otimes \left(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\right)^{-1} \right\} \mathbf{D}, \\
\widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,2} = & (4K)^{-1} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbf{D}^\top E \Bigg[\pi^{m}_{(\mathcal{I}_k)} \Big\{ \mathrm{vec}\left(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\right) - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \mathrm{vec}\left(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\right) \Big\} \\
& \times \left\{ \mathrm{vec}\left(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\right) \right\}^\top \Bigg] \mathbf{D}, \\
\widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,3} = & K^{-1} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbf{D}^\top E \left[\pi^{m}_{(\mathcal{I}_k)} \left\{ \left(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\right) \otimes \left(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\right)^{-1} \right\} \right] \mathbf{D}.
\end{aligned}
$$

Clearly, it can be proved that $\widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,1} \to 2^{-1}\mathbf{D}^\top \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right)\mathbf{D}$ as $N_k \to \infty$. Following the calculation for A.14 discussed in Appendix A.3, we have $\widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,2} - \widetilde{\mathbf{\Pi}}^{\mathbf{S}}_{33,3} \to -2^{-1}\mathbf{D}^\top \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right)\mathbf{D} + 4^{-1}\mathbf{D}^\top \mathrm{vec}\big(\mathbf{\Omega}^{-1}\big)\Big\{\mathrm{vec}\big(\mathbf{\Omega}^{-1}\big)\Big\}^\top \mathbf{D} - (4K)^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{m'=1}^{M} \mathbf{D}^\top E\Big[\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\Big\{\mathrm{vec}\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\Big\}\Big\{\mathrm{vec}\big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\big)\Big\}^\top\Big]\mathbf{D}$, as $N_k \to \infty$. With $\overline{\boldsymbol{X}}_k | \mathbb{B} \to^p \boldsymbol{\mu}_{kZ_k}$ as $N_K \to \infty$, we apply the Portmanteau Lemma to verify $E\Big[\pi^{m}_{(\mathcal{I}_k)}\pi^{m'}_{(\mathcal{I}_k)}\big\{\mathrm{vec}\big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\big)\big\}\big\{\mathrm{vec}\big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\big)\big\}^\top\Big] \to E\Big[\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)}\big\{\mathrm{vec}\big(\boldsymbol{b}^*_{km}\boldsymbol{b}^{*\top}_{km}\big)\big\}\big\{\mathrm{vec}\big(\boldsymbol{b}^*_{km'}\boldsymbol{b}^{*\top}_{km'}\big)\big\}^\top\Big]$. Thus, we can conclude

$$
\begin{aligned}
\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{33} \to^p & 4^{-1}\mathbf{D}^\top \mathrm{vec}\big(\mathbf{\Omega}^{-1}\big)\Big\{\mathrm{vec}\big(\mathbf{\Omega}^{-1}\big)\Big\}^\top \mathbf{D} \\
& - 4^{-1}\sum_{m=1}^{M}\sum_{m'=1}^{M} \mathbf{D}^\top E\big[\pi^{m*}_{(\mathcal{I}_k)}\pi^{m'*}_{(\mathcal{I}_k)} \left\{\mathrm{vec}\left(\boldsymbol{b}^*_{km}\boldsymbol{b}^{*\top}_{km}\right)\right\}\left\{\mathrm{vec}\left(\boldsymbol{b}^*_{km'}\boldsymbol{b}^{*\top}_{km'}\right)\right\}^\top\big]\mathbf{D} \\
= & -\mathbf{\Pi}_{33}.
\end{aligned}
$$

Similarly, based on the independence among $\overline{\boldsymbol{X}}_k$ from different classes and by applying Chebyshev's Law of Large Numbers, we can show that $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44} \xrightarrow{p} E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44})$. Subsequently, using the results from Appendix B.1 concerning the asymptotic form of $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44})$, we derive that

$$
\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{44} \to^p -2^{-1}\mathbf{D}^\top \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)\mathbf{D} = -\mathbf{\Pi}_{44}.
$$

Next, we consider the analysis of the off-diagonal block matrices. Since $\overline{\boldsymbol{X}}_k$s are independent, by applying similar operations as used for the on-diagonal matrices, we can prove that $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{i_1,i_2} \to^p E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{i_1,i_2})$, for $1 \le i_1 < i_2 \le 4$, as $K \to \infty$. Furthermore, with the detailed asymptotic form of $E(\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{i_1,i_2})$, we can obtain that $\widehat{\mathbf{\Pi}}^{\mathbf{S}}_{i_1,i_2} \to^p -\mathbf{\Pi}_{i_1,i_2}$, under Conditions (C1)-(C5).

$$
\begin{aligned}
\widehat{\mathbf{\Pi}}_{12,m} &= \sum_{k=1}^{K} \pi^{m}_{(\mathcal{I}_k)} \Big(\widetilde{\boldsymbol{\alpha}}_m - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \widetilde{\boldsymbol{\alpha}}_{m'}\Big) \boldsymbol{b}_{km}^\top, \\
\widehat{\mathbf{\Pi}}_{13} &= 2^{-1} \sum_{k=1}^{K}\sum_{m=1}^{M} \pi^{m}_{(\mathcal{I}_k)} \widetilde{\boldsymbol{\alpha}}_m \Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top \mathbf{D}, \\
\widehat{\mathbf{\Pi}}_{14} &= 2^{-1} \sum_{k=1}^{K}\sum_{m=1}^{M} \pi^{m}_{(\mathcal{I}_k)} \widetilde{\boldsymbol{\alpha}}_m \left\{\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\right\}^\top \mathbf{D}, \\
\widehat{\mathbf{\Pi}}_{23,m} &= -\sum_{k=1}^{K} \pi^{m}_{(\mathcal{I}_k)} \Big\{\boldsymbol{b}_{km}^\top \otimes (N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}\Big\}\mathbf{D} \\
&\quad + 2^{-1} \sum_{k=1}^{K} \pi^{m}_{(\mathcal{I}_k)} \boldsymbol{b}_{km} \Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \mathrm{vec}\Big(\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}^\top \mathbf{D}, \\
\widehat{\mathbf{\Pi}}_{24,m} &= -\sum_{k=1}^{K} \pi^{m}_{(\mathcal{I}_k)} \Big\{\boldsymbol{b}_{km}^\top \otimes N_k^{-1}(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega})^{-1}\Big\}\mathbf{D} \\
&\quad + 2^{-1} \sum_{k=1}^{K} \pi^{m}_{(\mathcal{I}_k)} \boldsymbol{b}_{km} \left\{\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\right\}^\top \mathbf{D}, \\
\widehat{\mathbf{\Pi}}_{34} &= 2^{-1} \sum_{k=1}^{K} \mathbf{D}^\top \Big\{N_k^{-1}\big(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\big)^{-1} \otimes \big(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\big)^{-1}\Big\}\mathbf{D} \\
&\quad + 4^{-1} \sum_{k=1}^{K}\sum_{m=1}^{M} \pi^{m}_{(\mathcal{I}_k)} \mathbf{D}^\top \Big\{\mathrm{vec}\Big(\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big)\Big\} \Big[\mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) - \sum_{m'=1}^{M} \pi^{m'}_{(\mathcal{I}_k)} \Big\{ \\
&\quad \mathrm{vec}\Big(N_k^{-1}\boldsymbol{b}_{km'}\boldsymbol{b}_{km'}^\top\Big)\Big\}\Big]^\top \mathbf{D} - \sum_{k=1}^{K}\sum_{m=1}^{M} \pi^{m}_{(\mathcal{I}_k)} \mathbf{D}^\top \Big\{\Big(N_k^{-1}\boldsymbol{b}_{km}\boldsymbol{b}_{km}^\top\Big) \otimes \big(N_k^{-1}\mathbf{\Sigma} + \mathbf{\Omega}\big)^{-1}\Big\}\mathbf{D}.
\end{aligned}
$$

Appendix B.3: Proof of Theorem 2

The theorem conclusion can be proved in two steps. In the first step, we will establish the consistency of $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$. In the second step, we study the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}$.

STEP 1. Note that the log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$ can be shown to be a strictly concave function in $\boldsymbol{\theta}$, as demonstrated in Appendix A. Therefore, if we can show that

$\mathcal{L}(\boldsymbol{\theta})$ has a consistent local maximizer, it must also be a consistent global maximizer. Recall that the form of $\boldsymbol{\theta}$ is given by $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \nu^\top, \text{vech}(\boldsymbol{\Omega})^\top, \text{vech}(\boldsymbol{\Sigma})^\top)^\top$. Next, we will show that $\mathcal{L}(\boldsymbol{\theta})$ has a local minimizer, where the components associated with $\boldsymbol{\Sigma}$ are $\sqrt{N}$-consistent, while the other components are $\sqrt{K}$-consistent. Leveraging the standardization matrix $\mathbf{S}$, and following the approach outlined in Fan and Li (2001), the existence of such a local minimizer is guaranteed by the fact that for any arbitrarily small $\varepsilon > 0$, there exists a sufficiently large constant C_ε such that

$$\liminf_{N\to\infty} P\left\{ \sup_{\|\boldsymbol{u}\|=1} \mathcal{L}\left(\boldsymbol{\theta} + C_\varepsilon \mathbf{S}\boldsymbol{u}\right) < \mathcal{L}\left(\boldsymbol{\theta}\right) \right\} \geq 1 - \varepsilon. \tag{B.1}$$

By the Taylor's expansion of the likelihood function, we have

$$\begin{aligned} \mathbf{D}_n(\boldsymbol{u}) &= \mathcal{L}\left(\boldsymbol{\theta} + C_\varepsilon \mathbf{S}\boldsymbol{u}\right) - \mathcal{L}\left(\boldsymbol{\theta}\right) \\ &= C_\varepsilon \left\{ \mathbf{S}\dot{\mathcal{L}}\left(\boldsymbol{\theta}\right) \right\}^\top \boldsymbol{u} - \frac{1}{2} C_\varepsilon^2 \boldsymbol{u}^\top \left\{ -\mathbf{S}\ddot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\mathbf{S} \right\} \boldsymbol{u} + o\left(\|\mathbf{S}^2\|_F\right). \end{aligned} \tag{B.2}$$

To begin with, we compute the first-order term of $\boldsymbol{u}$. Recall that it has been established in Appendix A and B.1 that $E\left\{\dot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\right\} = \mathbf{0}$ and $\mathbf{S}cov\left\{\dot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\right\}\mathbf{S} \to \boldsymbol{\Pi}$. By applying the central limit theorem, it follows that $\mathbf{S}\dot{\mathcal{L}}\left(\boldsymbol{\theta}\right) \to N(\mathbf{0}, \boldsymbol{\Pi})$ as $N \to \infty$. Consequently, with the Cauchy-Schwarz inequality, $\left\{\mathbf{S}\dot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\right\}^\top \boldsymbol{u}$ can be uniformly bounded by $\|\mathbf{S}\dot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\| = O_p(1)$. Next, we calculate the second-order term. As stated in Appendix B.2, $-\mathbf{S}\ddot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\mathbf{S} \to_p \boldsymbol{\Pi}$, which implies that $\lambda_{\min}\left\{-\mathbf{S}\ddot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\mathbf{S}\right\} \to_p \lambda_{\min}(\boldsymbol{\Pi})$, where $\lambda_{\min}(\mathbf{A})$ denotes the minimal eigenvalue of the matrix $\mathbf{A}$. Furthermore, we have the inequality $\boldsymbol{u}^\top\left\{-\mathbf{S}\ddot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\mathbf{S}\right\}\boldsymbol{u} \geq \lambda_{\min}\left\{-\mathbf{S}\ddot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\mathbf{S}\right\} = \lambda_{\min}(\boldsymbol{\Pi}) + o_p(1)$. As a consequence, we have $\boldsymbol{u}^\top\left\{-\mathbf{S}\ddot{\mathcal{L}}\left(\boldsymbol{\theta}\right)\mathbf{S}\right\}\boldsymbol{u} > 2^{-1}\lambda_{\min}(\boldsymbol{\Pi})$ with probability tending to one. Thus, for the ε mentioned above, we can take a sufficiently large C_ε such that $\sup_{\|\boldsymbol{u}\|=1} \mathcal{L}\left(\boldsymbol{\theta} + C_\varepsilon \mathbf{S}\boldsymbol{u}\right) < \mathcal{L}\left(\boldsymbol{\theta}\right)$. At this stage, the proof of consistency has been completed. Additionally, using the specific form of the standardization matrix $\mathbf{S}$, we can determine the convergence rates for each component.

STEP 2. By applying the Taylor's expansion to the gradient condition as $\dot{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\text{MLE}}) =$

$\mathbf{0}$, we obtain

$$\mathbf{0} = \mathbf{S}\dot{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\text{MLE}}) = \mathbf{S}\dot{\mathcal{L}}(\boldsymbol{\theta}) + \mathbf{S}\ddot{\mathcal{L}}(\boldsymbol{\theta})\mathbf{S}\Big\{\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta})\Big\} + o_p\Big(\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta})\Big).$$

Note that under Conditions (C1)-(C5), it has been established in Appendix B.2 that $\widehat{\mathbf{\Pi}}^{\mathbf{S}} = \mathbf{S}\ddot{\mathcal{L}}\,(\boldsymbol{\theta})\,\mathbf{S} \to_p -\mathbf{\Pi}$ as $N \to \infty$. Therefore, we may replace the term $\mathbf{S}\ddot{\mathcal{L}}\,(\boldsymbol{\theta})$ with $\{-\mathbf{\Pi} + o_p(1)\}$. The equation above can then be rewritten as

$$\begin{aligned}\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}) &= -\Big\{\mathbf{S}\ddot{\mathcal{L}}(\boldsymbol{\theta})\mathbf{S}\Big\}^{-1}\mathbf{S}\dot{\mathcal{L}}(\boldsymbol{\theta}) + o_p\Big(\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta})\Big) \\ &= \mathbf{\Pi}^{-1}\mathbf{S}\dot{\mathcal{L}}(\boldsymbol{\theta}) + o_p\Big(\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{MLE}})\Big).\end{aligned}$$

As verified in Step 1, $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$ is a consistent estimator with a divergence rate of order $O_p(\text{vecdiag}(\boldsymbol{S}^{-1}))$, where $\text{vecdiag}(\boldsymbol{S}^{-1})$ denotes the vector formed by the on-diagonal elements of the diagonal matrix $\boldsymbol{S}^{-1}$. Furthermore, it is also shown in Step 1 that $\mathbf{S}\dot{\mathcal{L}}\,(\boldsymbol{\theta}) \to N(\mathbf{0}, \mathbf{\Pi})$. Applying Slutsky's theorem, we then conclude

$$\mathbf{S}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}) \to^d N\big(\mathbf{0}, \mathbf{\Pi}^{-1}\big).$$

This completes the derivation of the asymptotic properties of $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$ in Appendix B.3.

APPENDIX C

Appendix C.1: A Tensor-based EM Algorithm

The E-Step

Recall that $\mathbf{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_N)^\top \in \mathbb{R}^{N\times p}$ is the design matrix. Next, define a one-hot type response matrix $\mathbf{Y} = (Y_{ik}) \in \mathbb{R}^{N\times K}$ where $Y_{ik} = I(Y_i = k)$. Note that $\mathbf{Y}$ is of an extremely sparse structure. Therefore, it can be stored in a sparse matrix form to save memory. Define $\boldsymbol{Y} = \mathbf{Y}^\top\mathbf{1}_N \in \mathbb{R}^K$ and $\mathbf{1}_N = (1, \ldots, 1)^\top \in \mathbb{R}^N$. We then have $\overline{\mathbf{X}} = [(\mathbf{X}^\top\mathbf{Y})\{\text{diag}(\boldsymbol{Y})\}^{-1}]^\top = (\overline{\boldsymbol{X}}_1, \ldots, \overline{\boldsymbol{X}}_K)^\top \in \mathbb{R}^{K\times p}$, where $\text{diag}(\boldsymbol{A})$

denotes the diagonal matrix formed from the vector $\boldsymbol{A}$. We next consider how to execute the expectation step (3) in a computationally more efficient way. Let $\widehat{\boldsymbol{\theta}}^{(t)} = (\widehat{\alpha}_1^{(t)}, \ldots, \widehat{\alpha}_{M-1}^{(t)}, \mathrm{vech}(\widehat{\boldsymbol{\Sigma}}^{(t)})^\top, \mathrm{vech}(\widehat{\boldsymbol{\Omega}}^{(t)})^\top, \widehat{\boldsymbol{\nu}}_1^{(t)\top}, \ldots, \widehat{\boldsymbol{\nu}}_M^{(t)\top})^\top \in \mathbb{R}^q$ be the parameter obtained in the tth step. In the $(t+1)$th step, we then should update the posterior probability according to equation (4) as

$$\begin{aligned}\pi_{(\mathcal{I}_k)}^{m(t+1)} =& \Big[\widehat{\alpha}_m^{(t)} \exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_m^{(t)}\big)^\top \big(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)} + \widehat{\boldsymbol{\Omega}}^{(t)}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_m^{(t)}\big)\Big\}\Big] \\ &\times \Big[\sum_{m'=1}^{M} \widehat{\alpha}_{m'}^{(t)} \exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\big)^\top \big(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)} + \widehat{\boldsymbol{\Omega}}^{(t)}\big)^{-1}\big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\big)\Big\}\Big]^{-1}.\end{aligned} \tag{C.1}$$

To this end, we find that the high dimensional matrices $(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)} + \widehat{\boldsymbol{\Omega}}^{(t)})$ needs to be inverted for a total of K times in (C.1).

This is obviously computationally inefficient. To fix the problem, we re-write $d_{km}^{(t)} = (\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)})^\top (N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)} + \widehat{\boldsymbol{\Omega}}^{(t)})^{-1}(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)})$ as

$$\begin{aligned}d_{km}^{(t)} =& \Big\{\Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{-1/2}\Big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\Big)\Big\}^\top \Big\{N_k^{-1}\mathbf{I} + \Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{-1/2}\widehat{\boldsymbol{\Omega}}^{(t)}\Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{-1/2}\Big\}^{-1} \\ &\times \Big\{\Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{-1/2}\Big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\Big)\Big\} \\ =& \Big\{\Big(\mathbf{U}^{(t)}\Big)^\top\Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{-1/2}\Big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\Big)\Big\}^\top \Big\{N_k^{-1}\mathbf{I} + \mathrm{diag}\big(\boldsymbol{\lambda}^{(t)}\big)\Big\}^{-1} \\ &\times \Big\{\Big(\mathbf{U}^{(t)}\Big)^\top\Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{-1/2}\Big(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\Big)\Big\},\end{aligned} \tag{C.2}$$

where $(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2}\widehat{\boldsymbol{\Omega}}^{(t)}(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2} = \mathbf{U}^{(t)}\mathrm{diag}(\boldsymbol{\lambda}^{(t)})(\mathbf{U}^{(t)})^\top$ with $\boldsymbol{\lambda}^{(t)} = (\lambda_j^{(t)}) \in \mathbb{R}^p$ and $\mathbf{U}^{(t)} = (\boldsymbol{U}_1^{(t)}, \ldots, \boldsymbol{U}_p^{(t)}) \in \mathbb{R}^{p\times p}$ being an orthogonal matrix. Note that both the matrices $(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2}$ and the eigenvalue decomposition $\{\mathbf{U}^{(t)}, \boldsymbol{\lambda}^{(t)}\}$ do not depend on the class label k. Therefore, once they are computed, they can be repeatedly used for every $1 \le k \le K$ in the $(t+1)$th step. That leads to a significant reduction in computational cost. More specifically, define $\overline{\boldsymbol{X}}_k^{*(t)} = (\mathbf{U}^{(t)})^\top(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2}\overline{\boldsymbol{X}}_k$ and $\widehat{\boldsymbol{\nu}}_m^{*(t)} = (\mathbf{U}^{(t)})^\top(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2}\widehat{\boldsymbol{\nu}}_m^{(t)}$. We next define two 3-dimensional tensors as $\mathcal{D}^{(t)} = (\boldsymbol{D}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p}$ with $\boldsymbol{D}_{km}^{(t)} = \overline{\boldsymbol{X}}_k^{(t)} - \widehat{\boldsymbol{\nu}}_m^{(t)} \in \mathbb{R}^p$ and $\mathcal{D}^{*(t)} = (\boldsymbol{D}_{km}^{*(t)} :$

$1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p}$ with $\boldsymbol{D}_{km}^{*(t)} = \overline{\boldsymbol{X}}_k^{*(t)} - \widehat{\boldsymbol{\nu}}_m^{*(t)} \in \mathbb{R}^p$. Define another 3-dimensional tensor as $\mathcal{G}^{(t)} = (\boldsymbol{G}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p}$ with $\boldsymbol{G}_{km}^{(t)} = \{(N_k^{-1} + \lambda_j^{(t)})^{-1} : 1 \le j \le p\} \in \mathbb{R}^p$. We then can obtain another 3-dimensional tensor as $\mathcal{H}^{(t)} = (\boldsymbol{H}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) = \mathcal{D}^{*(t)} \odot \mathcal{G}^{(t)} \odot \mathcal{D}^{*(t)} \in \mathbb{R}^{K\times M\times p}$, where $\odot$ stands for the component-wise product. It follows then $d_{km}^{(t)} = \mathbf{1}_N^\top \boldsymbol{H}_{km}^{(t)}$ in (C.2).

The M-Step

Next, we consider rewriting the algorithm of the M-step into a tensor-based format. Recall that $\widehat{\mathbf{\Sigma}}^{(t)}$, $\widehat{\mathbf{\Omega}}^{(t)}$, $\widehat{\alpha}_m^{(t)}$ and $\widehat{\boldsymbol{\nu}}_m^{(t)}$ with $1 \le m \le M$ are the estimators obtained in the tth step. We begin with computing $\widehat{\alpha}_m^{(t+1)}$ for $1 \le m \le M$. According to equation (3), $\widehat{\alpha}_m^{(t+1)} = K^{-1}(\sum_{k=1}^K \pi_{(\mathcal{I}_k)}^m)$ with $1 \le m \le M$. Define $\mathbf{H}^{(t)} = (\pi_{(\mathcal{I}_k)}^{m(t+1)}) \in \mathbb{R}^{K\times M}$ as the matrix of posterior possibilities obtained in the $(t+1)$th step. Then, we can calculate $(\widehat{\alpha}_1^{(t+1)}, \dots, \widehat{\alpha}_M^{(t+1)}) = K^{-1}\mathbf{1}_K^\top \mathbf{H}^{(t)}$. We next consider how to compute $\widehat{\mathbf{\Sigma}}^{(t+1)}$. By equation (3), the $\widehat{\mathbf{\Sigma}}^{(t+1)}$ can be decomposed and calculated as

$$\widehat{\mathbf{\Sigma}}^{(t+1)} = (N-K)^{-1}\Bigg(\sum_{k=1}^K\sum_{m=1}^M \pi_{(\mathcal{I}_k)}^{m(t+1)}\Bigg[\sum_{i\in\mathcal{I}_k}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k\right)\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k\right)^\top +$$

$$N_k^{-1}\Big\{\widehat{\mathbf{\Sigma}}^{(t)}\left(N_k^{-1}\widehat{\mathbf{\Sigma}}^{(t)} + \widehat{\mathbf{\Omega}}^{(t)}\right)^{-1}\left(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_m^{(t)}\right)\left(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_m^{(t)}\right)^\top\left(N_k^{-1}\widehat{\mathbf{\Sigma}}^{(t)} + \widehat{\mathbf{\Omega}}^{(t)}\right)^{-1}\widehat{\mathbf{\Sigma}}^{(t)}$$

$$-\widehat{\mathbf{\Sigma}}^{(t)}\left(N_k^{-1}\widehat{\mathbf{\Sigma}}^{(t)} + \widehat{\mathbf{\Omega}}^{(t)}\right)^{-1}\widehat{\mathbf{\Sigma}}^{(t)}\Big\}\Bigg]\Bigg) = (N-K)^{-1}\left(\mathbf{Q}_1 + \mathbf{Q}_2\right),$$

$$\text{where } \mathbf{Q}_1 = \sum_{i=1}^N\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{Y_i}\right)\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{Y_i}\right)^\top,$$

$$\text{and } \mathbf{Q}_2 = \sum_{k=1}^K\sum_{m=1}^M N_k^{-1}\pi_{(\mathcal{I}_k)}^{m(t+1)}\widehat{\mathbf{\Sigma}}^{(t)}\Big\{\left(N_k^{-1}\widehat{\mathbf{\Sigma}}^{(t)} + \widehat{\mathbf{\Omega}}^{(t)}\right)^{-1}\left(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_m^{(t)}\right)\left(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_m^{(t)}\right)^\top \times$$

$$\left(N_k^{-1}\widehat{\mathbf{\Sigma}}^{(t)} + \widehat{\mathbf{\Omega}}^{(t)}\right)^{-1} - \left(N_k^{-1}\widehat{\mathbf{\Sigma}}^{(t)} + \widehat{\mathbf{\Omega}}^{(t)}\right)^{-1}\Big\}\widehat{\mathbf{\Sigma}}^{(t)},$$

We should start with $\mathbf{Q}_1 = \sum_{i=1}^N(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{Y_i})(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{Y_i})^\top = (\mathbf{X} - \mathbf{Y}\overline{\mathbf{X}})^\top(\mathbf{X} - \mathbf{Y}\overline{\mathbf{X}})$. We next consider $\mathbf{Q}_2$. Recall that $(\widehat{\mathbf{\Sigma}}^{(t)})^{-1/2}\widehat{\mathbf{\Omega}}^{(t)}(\widehat{\mathbf{\Sigma}}^{(t)})^{-1/2} = \mathbf{U}^{(t)}\mathrm{diag}(\boldsymbol{\lambda}^{(t)})(\mathbf{U}^{(t)})^\top$. Then, the computation of $\mathbf{Q}_2$ can be simplified as $\mathbf{Q}_2 = \sum_{k=1}^K \sum_{m=1}^M N_k^{-1} \pi_{(\mathcal{I}_k)}^{m(t+1)} \mathbf{Q}_{2km}$ with

$$\begin{aligned}\mathbf{Q}_{2km} = \left(\widetilde{\boldsymbol{\Sigma}}^{(t)}\right)^\top \Bigg[& \left\{ N_k^{-1}\mathbf{I} + \operatorname{diag}\left(\boldsymbol{\lambda}^{(t)}\right) \right\}^{-1} \left\{ \left(\mathbf{U}^{(t)}\right)^\top \left(\widehat{\boldsymbol{\Sigma}}^{(t)}\right)^{-1/2} \left(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\right) \right\} \\ & \times \left\{ \left(\mathbf{U}^{(t)}\right)^\top \left(\widehat{\boldsymbol{\Sigma}}^{(t)}\right)^{-1/2} \left(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)}\right) \right\}^\top \left\{ N_k^{-1}\mathbf{I} + \operatorname{diag}\left(\boldsymbol{\lambda}^{(t)}\right) \right\}^{-1} \\ & - \left\{ N_k^{-1}\mathbf{I} + \operatorname{diag}\left(\boldsymbol{\lambda}^{(t)}\right) \right\}^{-1} \Bigg] \widetilde{\boldsymbol{\Sigma}}^{(t)},\end{aligned}$$

where $\widetilde{\boldsymbol{\Sigma}}^{(t)} = (\mathbf{U}^{(t)})^\top (\widehat{\boldsymbol{\Sigma}}^{(t)})^{1/2} \in \mathbb{R}^{p\times p}$. Since $\widehat{\boldsymbol{\Sigma}}^{(t)}$ and $\boldsymbol{\lambda}^{(t)}$ are fixed in the $(t+1)$th step, the matrix $\widetilde{\boldsymbol{\Sigma}}^{(t)}$ can be repeatedly used in M-step once they are computed.

Recall that $\mathcal{D}^{*(t)} = (\boldsymbol{D}_{km}^{*(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p}$ with $\boldsymbol{D}_{km}^{*(t)} = \overline{\boldsymbol{X}}_k^{*(t)} - \widehat{\boldsymbol{\nu}}_m^{*(t)} \in \mathbb{R}^p$ and $\mathcal{G}^{(t)} = (\boldsymbol{G}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p}$ with $\boldsymbol{G}_{km}^{(t)} = \{(N_k^{-1} + \lambda_j^{(t)})^{-1} : 1 \le j \le p\} \in \mathbb{R}^p$. We then can obtain another 3-dimensional tensor as $\mathcal{J}^{(t)} = (\boldsymbol{J}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) = \mathcal{G}^{(t)} \odot \mathcal{D}^{*(t)} \in \mathbb{R}^{K\times M\times p}$, where $\boldsymbol{J}_{km}^{(t)} = \{N_k^{-1}\mathbf{I} + \operatorname{diag}(\boldsymbol{\lambda})^{(t)}\}^{-1}\{(\mathbf{U}^{(t)})^\top(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2}(\overline{\boldsymbol{X}}_k - \widehat{\boldsymbol{\nu}}_{m'}^{(t)})\} \in \mathbb{R}^p$. Next, we consider expanding $\mathcal{J}^{(t)}$ into 4-dimensional tensors $\widetilde{\mathcal{J}}^{(t)} = (\widetilde{\mathbf{J}}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p\times p}$ and $\mathcal{G}^{(t)}$ into 4-dimensional tensors $\widetilde{\mathcal{G}}^{(t)} = (\widetilde{\mathbf{G}}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p\times p}$, where $\widetilde{\mathbf{J}}_{km}^{(t)} \in \mathbb{R}^{p\times p} = \boldsymbol{J}_{km}^{(t)}(\boldsymbol{J}_{km}^{(t)})^\top$ and $\widetilde{\mathbf{G}}_{km}^{(t)} = \operatorname{diag}(\boldsymbol{G}_{km}^{(t)}) \in \mathbb{R}^{p\times p}$. We define another 4-dimensional tensor as $\mathcal{K}^{(t)} = (\mathbf{K}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p\times p}$, where $\mathbf{K}_{km}^{(t)} = \widetilde{\mathbf{J}}_{km}^{(t)} - \widetilde{\mathbf{G}}_{km}^{(t)} \in \mathbb{R}^{p\times p}$. For simplicity, define a 4-dimensional tensor as $\mathcal{P}^{*(t)} = (\mathcal{P}_{km}^{*(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p\times p}$, where $\mathcal{P}_{km}^{*(t)} = \mathbf{1}_p\mathbf{1}_p^\top N_k^{-1}\pi_{(\mathcal{I}_k)}^{m(t+1)} \in \mathbb{R}^{p\times p}$. Then we can obtain a new 4-dimensional tensor as $\widetilde{\mathcal{K}}^{(t)} = (\widetilde{\mathbf{K}}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) = \mathcal{K}^{(t)} \odot \mathcal{P}^{*(t)} \in \mathbb{R}^{K\times M\times p\times p}$. By summarizing the tensor $\widetilde{\mathcal{K}}^{(t)}$ along the first and second dimensions, we can eventually obtain a matrix $\mathbf{K}^{(t)} = \sum_{k=1}^K \sum_{m=1}^M \widetilde{\mathbf{K}}_{km}^{(t)} \in \mathbb{R}^{p\times p}$ with straightforward operation as flattening and dimension transformations. Then, $\mathbf{Q}_2 = (\widetilde{\boldsymbol{\Sigma}}^{(t)})^\top \mathbf{K}^{(t)} \widetilde{\boldsymbol{\Sigma}}^{(t)}$. Therefore, we have completed the rewriting of $\widehat{\boldsymbol{\Sigma}}^{(t+1)}$.

Then, we solve the problem of computing $\widehat{\boldsymbol{\Omega}}^{(t+1)}$. In the $(t+1)$th step, we then should update the covariance according to equation (3) as

$$\begin{aligned}\widehat{\boldsymbol{\Omega}}^{(t+1)} =& K^{-1}\Bigg\{\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m(t+1)}\widehat{\boldsymbol{\Omega}}^{(t)}\left(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)}+\widehat{\boldsymbol{\Omega}}^{(t)}\right)^{-1}\left(\overline{\boldsymbol{X}}_k-\widehat{\nu}_m^{(t)}\right)\left(\overline{\boldsymbol{X}}_k-\widehat{\nu}_m^{(t)}\right)^\top \\ &\times\left(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)}+\widehat{\boldsymbol{\Omega}}^{(t)}\right)^{-1}\widehat{\boldsymbol{\Omega}}^{(t)}+\sum_{k=1}^{K}\widehat{\boldsymbol{\Omega}}^{(t)}\left(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)}+\widehat{\boldsymbol{\Omega}}^{(t)}\right)^{-1}N_K^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)}\Bigg\}. \quad \text{(C.3)}\end{aligned}$$

By the similar arguments, we can obtain that $\widehat{\boldsymbol{\Omega}}^{(t+1)} = K^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}\pi_{(\mathcal{I}_k)}^{m(t+1)}\mathbf{Q}_{km}$, where $\mathbf{Q}_{km}$ is given as

$$\begin{aligned}\mathbf{Q}_{km} =& \left(\widetilde{\boldsymbol{\Omega}}^{(t)}\right)^\top\Bigg[\left\{N_k^{-1}\mathbf{I}+\mathrm{diag}\big(\boldsymbol{\lambda}^{(t)}\big)\right\}^{-1}\left\{\left(\mathbf{U}^{(t)}\right)^\top\left(\widehat{\boldsymbol{\Sigma}}^{(t)}\right)^{-1/2}\left(\overline{\boldsymbol{X}}_k-\widehat{\boldsymbol{\nu}}_{m'}^{(t)}\right)\right\} \\ &\times\left\{\left(\mathbf{U}^{(t)}\right)^\top\left(\widehat{\boldsymbol{\Sigma}}^{(t)}\right)^{-1/2}\left(\overline{\boldsymbol{X}}_k-\widehat{\boldsymbol{\nu}}_{m'}^{(t)}\right)\right\}^\top\left\{N_k^{-1}\mathbf{I}+\mathrm{diag}\big(\boldsymbol{\lambda}^{(t)}\big)\right\}^{-1}\widetilde{\boldsymbol{\Omega}}^{(t)} \\ &+\left\{N_k^{-1}\mathbf{I}+\mathrm{diag}\big(\boldsymbol{\lambda}^{(t)}\big)\right\}^{-1}N_K^{-1}\widetilde{\boldsymbol{\Sigma}}^{(t)}\Bigg],\end{aligned}$$

where $\widetilde{\boldsymbol{\Omega}}^{(t)} = (\mathbf{U}^{(t)})^\top(\widehat{\boldsymbol{\Sigma}}^{(t)})^{-1/2}\widehat{\boldsymbol{\Omega}}^{(t)} \in \mathbb{R}^{p\times p}$. Since $\widehat{\boldsymbol{\Omega}}^{(t)}$ and $\boldsymbol{\lambda}^{(t)}$ are fixed in the $(t+1)$th step, the matrix $\widetilde{\boldsymbol{\Omega}}^{(t)}$ can be repeatedly used in M-step once they are computed. With the tensors $\widetilde{\mathcal{G}}^{(t)} \in \mathbb{R}^{K\times M\times p\times p}$ and $\mathcal{P}^{*(t)} \in \mathbb{R}^{K\times M\times p\times p}$ defined before, we can obtain a 4-dimensional tensor as $\mathcal{M}^{(t)} = (\mathbf{M}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) = \widetilde{\mathcal{G}}^{(t)} \odot \mathcal{P}^{*(t)} \in \mathbb{R}^{K\times M\times p\times p}$. For simplicity, define a 4-dimensional tensor as $\mathcal{P}^{(t)} = (\mathbf{P}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) \in \mathbb{R}^{K\times M\times p\times p}$, where $\mathbf{P}_{km}^{(t)} = \mathbf{1}_p\mathbf{1}_p^\top\pi_{(\mathcal{I}_k)}^{m(t+1)} \in \mathbb{R}^{p\times p}$. Then, we can similarly obtain two new 4-dimensional tensors as $\widehat{\mathcal{K}}^{(t)} = (\widehat{\mathbf{K}}_{km}^{(t)} : 1 \le k \le K, 1 \le m \le M) = \widetilde{\mathcal{J}}^{(t)} \odot \mathcal{P}^{(t)} \in \mathbb{R}^{K\times M\times p\times p}$. By summarizing the tensors along the first and second dimensions, we can eventually obtain matrices $\mathbf{M}^{(t)} = \sum_{k=1}^{K}\sum_{m=1}^{M}\mathbf{M}_{km}^{(t)} \in \mathbb{R}^{p\times p}$ and $\widehat{\mathbf{K}}^{(t)} = \sum_{k=1}^{K}\sum_{m=1}^{M}\widehat{\mathbf{K}}_{km}^{(t)} \in \mathbb{R}^{p\times p}$ with straightforward operation as flattening and dimension transformations. Then, $\widehat{\boldsymbol{\Omega}}^{(t+1)} = K^{-1}\{(\widetilde{\boldsymbol{\Omega}}^{(t)})^\top\widehat{\mathbf{K}}^{(t)}\widetilde{\boldsymbol{\Omega}}^{(t)} + (\widetilde{\boldsymbol{\Omega}}^{(t)})^\top\mathbf{M}^{(t)}\widetilde{\boldsymbol{\Sigma}}^{(t)}\}$.

Finally, we consider how to estimate $\widehat{\boldsymbol{\nu}}_m^{(t+1)}$ according to equation (3). We have

$$\widehat{\boldsymbol{\nu}}_m^{(t+1)} = \left\{\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\left(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)}+\widehat{\boldsymbol{\Omega}}^{(t)}\right)^{-1}\right\}^{-1}\left\{\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\left(N_k^{-1}\widehat{\boldsymbol{\Sigma}}^{(t)}+\widehat{\boldsymbol{\Omega}}^{(t)}\right)^{-1}\overline{\boldsymbol{X}}_k\right\},$$

$$
\begin{aligned}
&=\Big(\widehat{\boldsymbol{\Sigma}}^{(t)}\Big)^{1/2}\mathbf{U}^{(t)}\Bigg[\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\Big\{N_k^{-1}\mathbf{I}+\mathrm{diag}\big(\boldsymbol{\lambda}^{(t)}\big)\Big\}^{-1}\Bigg]^{-1} \\
&\quad\times\Bigg[\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\Big\{N_k^{-1}\mathbf{I}+\mathrm{diag}\big(\boldsymbol{\lambda}^{(t)}\big)\Big\}^{-1}\overline{\boldsymbol{X}}_k^{*(t)}\Bigg].
\end{aligned}
$$

To compute $[\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\{N_k^{-1}\mathbf{I}+\mathrm{diag}(\boldsymbol{\lambda}^{(t)})\}^{-1}]^{-1}$, we define a 3-dimensional tensor as $\widetilde{\mathcal{P}}^{(t)}=(\widetilde{\boldsymbol{P}}_{km}^{(t)}:1\le k\le K,1\le m\le M)\in\mathbb{R}^{K\times M\times p}$ with $\widetilde{\boldsymbol{P}}_{km}^{(t)}=\pi_{(\mathcal{I}_k)}^{m(t+1)}\mathbf{1}_p\in\mathbb{R}^p$. Then, we can obtain a 3-dimensional tensor $\mathcal{R}^{(t)}=\widetilde{\mathcal{P}}^{(t)}\odot\mathcal{D}^{(t)}=(\mathbf{R}_k^{(t)}:1\le k\le K)\in\mathbb{R}^{K\times M\times p}$. By summarizing the tensor $\mathcal{R}^{(t)}$ along the first dimension, we can obtain a matrix $\mathbf{R}^{(t)}=\sum_{k=1}^{K}\mathbf{R}_k^{(t)}\in\mathbb{R}^{M\times p}$ with straightforward operation as flattening and dimension transformations. Therefore, we have $[\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\{N_k^{-1}\mathbf{I}+\mathrm{diag}(\boldsymbol{\lambda}^{(t)})\}^{-1}]^{-1}=(R^{(t)})^{-1}$. Similarly, we can compute the term $\sum_{k=1}^{K}\pi_{(\mathcal{I}_k)}^{m(t+1)}\{N_k^{-1}\mathbf{I}+\mathrm{diag}(\boldsymbol{\lambda}^{(t)})\}^{-1}\overline{\boldsymbol{X}}_k^{*(t)}$. Define a 3-dimensional tensor $\widetilde{\mathcal{X}}^{(t)}=(\widetilde{\boldsymbol{X}}_{km}^{(t)}:1\le k\le K,1\le m\le M)\in\mathbb{R}^{K\times M\times p}$ with $\widetilde{\boldsymbol{X}}_{km}^{(t)}=\overline{\boldsymbol{X}}_k^{*(t)}\in\mathbb{R}^p$. Then, we can have a new 3-dimensional tensor $\mathcal{Q}^{(t)}=\mathcal{R}^{(t)}\odot\widetilde{\mathcal{X}}^{(t)}=(\mathbf{Q}_k^{(t)}:1\le k\le K)\in\mathbb{R}^{K\times M\times p}$. Summarizing the tensor $\mathcal{Q}^{(t)}$ along the first dimension, we can obtain a matrix $\mathbf{Q}^{(t)}=\sum_{k=1}^{K}\mathbf{Q}_k^{(t)}\in\mathbb{R}^{M\times p}$. Therefore, we have $(\widehat{\boldsymbol{\nu}}_1^{(t+1)},\ldots,\widehat{\boldsymbol{\nu}}_M^{(t+1)})^\top=(\widehat{\boldsymbol{\Sigma}}^{(t)})^{1/2}\mathbf{U}^{(t)}(\mathbf{R}^{(t)})^{-1}\mathbf{Q}^{(t)}$. The construction of the tensor-based EM algorithm has been completed above.

Appendix C.2: The Log-likelihood Function

To derive the tensor formulation of the log-likelihood function, we proceed to decompose $\mathcal{L}\big(\boldsymbol{\theta}\big)$, which is given in the equation (2), into three parts.

$$
\begin{aligned}
\mathcal{L}_1\big(\boldsymbol{\theta}\big)&=-2^{-1}N\log\big|\boldsymbol{\Sigma}\big|-2^{-1}K\log\big|\boldsymbol{\Omega}\big|-\sum_{k=1}^{K}2^{-1}\log\big|N_k\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1}\big|,\\
\mathcal{L}_2\big(\boldsymbol{\theta}\big)&=\sum_{k=1}^{K}\log\Bigg[\sum_{m=1}^{M}\alpha_m\exp\Big\{-2^{-1}\big(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m\big)^\top\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m\big)\Big\}\Bigg],\\
\mathcal{L}_3\big(\boldsymbol{\theta}\big)&=\sum_{i=k}^{K}\sum_{i\in\mathcal{I}_k}-2^{-1}\big(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k\big)^\top\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{X}_i-\overline{\boldsymbol{X}}_k\big).
\end{aligned}
\tag{C.4}
$$

Next, we start with the first component $\mathcal{L}_1\big(\boldsymbol{\theta}\big)$.

$$\begin{aligned}\mathcal{L}_1\big(\boldsymbol{\theta}\big) &= -2^{-1}\sum_{k=1}^{K}\Big\{N_k\log\big|\boldsymbol{\Sigma}\big| + \log\big|\boldsymbol{\Omega}\big| + \log\big|N_k\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1}\big|\Big\}\\ &= -2^{-1}\sum_{k=1}^{K}\Big\{(N_k-1)\log\big|\boldsymbol{\Sigma}\big| + \log\big|N_k\boldsymbol{\Omega}+\boldsymbol{\Sigma}\big|\Big\}\\ &= -2^{-1}\sum_{k=1}^{K}\Big\{N_k\log\big|\boldsymbol{\Sigma}\big| + p\log(N_k) + \log\big|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1/2}+N_k^{-1}\mathbf{I}_{p\times p}\big|\Big\}.\end{aligned}$$

Recall that we have $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1/2} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$ with $\boldsymbol{\lambda} = (\lambda_j) \in \mathbb{R}^p$ and $\mathbf{U} = (\boldsymbol{U}_1,\ldots,\boldsymbol{U}_p) \in \mathbb{R}^{p\times p}$ being an orthogonal matrix. We can calculate $\log\big|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1/2}+N_k^{-1}\mathbf{I}_{p\times p}\big| = \sum_{j=1}^{p}\log(\lambda_j+N_K^{-1})$. Therefore, we have

$$\mathcal{L}_1\big(\boldsymbol{\theta}\big) = -2^{-1}\sum_{k=1}^{K}\Big\{N_k\log\big|\boldsymbol{\Sigma}\big| + p\log(N_k) + \sum_{j=1}^{p}\log(\lambda_j+N_K^{-1})\Big\}. \tag{C.5}$$

Then, we can deal with the second component $\mathcal{L}_2\big(\boldsymbol{\theta}\big)$. Obviously, we have computed $\big(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m\big)^\top\big(N_k^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Omega}\big)^{-1}\big(\overline{\boldsymbol{X}}_k-\boldsymbol{\nu}_m\big)$ as d_{km} in the E step. By the similar arguments, we can obtain its implementation. Lastly, we need to compute $\mathcal{L}_3\big(\boldsymbol{\theta}\big)$. Recall that the design matrix $\mathbf{X} \in \mathbb{R}^{N\times p}$, the one-hot type response matrix $\mathbf{Y} \in \mathbb{R}^{N\times K}$ and the class mean martrix $\overline{\mathbf{X}} = (\overline{\boldsymbol{X}}_1,\ldots,\overline{\boldsymbol{X}}_K)^\top \in \mathbb{R}^{K\times p}$ have been defined. We can obtain $\mathcal{L}_3\big(\boldsymbol{\theta}\big) = -2^{-1}\mathrm{tr}\{(\mathbf{X}-\mathbf{Y}\overline{\mathbf{X}})\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{Y}\overline{\mathbf{X}})^\top\}$.

APPENDIX D: Group Structures of the Real Dataset

Label	Group	Categories
1	Electric Kettles & Cookware	Electric Hot Plate, Automatic Herbal Medicine Pot, Wellness Kettle, Electric Kettle, Automatic Electric Tea Set, Electric Slow Cooker
2	Small Kitchen Appliances	Soy Milk Maker, Hand Mixer, Hand Mixer, Electric Meat Grinder, Panini Press, Pasta Maker, Juicer, Coffee Maker, Egg Cooker, Food Processor, High-Speed Blender, Toaster Oven, Pancake Maker
3	Large Kitchen Appliances	Multi-function Electric Pot, Microwave Oven, Electric Pressure Cooker, Air Fryer, Electric Lunch Box, Food Warmer Plate, Electric Oven, Induction Cooktop, Home Food Warmer Plate, Electric Oven, Rice Cooker, Combination Steam Oven
4	Laundry and Garment Care Tools	Washing Machine, Dryer, Washer-Dryer Combo, Steam Iron, Sanitizing Washer Dryer
5	Air Conditioning and Ventilation Systems	Central Air Conditioning, Air Conditioner, Portable Air Conditioner, Air Cooler, Fresh Air System, Electric Fan
6	Cleaning Appliances	Water Purifier & Air Purifier Filters, Air Purifier, Household Water Purifier, Robotic Vacuum Cleaner, Dehumidifier, Fabric Shaver, Household Dust Mite Eliminator, Electric Cleaning & Steam Equipment, Vacuum Cleaner, Humidifier
7	Other Home Appliances	Wine Cooler, Freezer, Smart TV, Refrigerator, Sterilizing Cabinet, Range Hood & Gas Stove Set, Dishwasher, Integrated Stove, Gas Stove, Solar Water Heater, Electric Water Heater, Air Source Heat Pump Water Heater, Gas Water Heater, Kitchen Appliance Accessories, Electric Heating Table, Heater, Desktop Telephone, Home Electronics, Water Dispenser, Tea Bar Machine, Smart Home Devices
8	Personal Care Appliances	Shaver Accessories, Electric Toothbrush Replacement Heads, Hair Curling Iron/Straightener, Electric Oral Irrigator, Personal Grooming Hair Removal & Trimmer, Hair Dryer, Electric Shaver, Electric Toothbrush, Beauty Devices, Electronic Scale, Hair Cutting Tools
9	Massage Equipment	Massager, Foot Massager, Home Massage & Physical Therapy Equipment, Massage Chair, Foot Bath, Home Electric Physical Therapy Device
10	Seafood	Softshell Turtles, Canned Prepared Seafood, Other Seafood, Crabs, Seafood Related Products, Seafood Gift Boxes, Seaweed Products, Shrimp & Prawns, Dried Sea Cucumbers, Seafood, Dried Seafood
11	Refrigerated & Frozen Foods	Pizza, Ready-to-Cook Baked Goods, Dumplings, Portuguese Egg Tarts, Birthday Cakes, Frozen Foods, Juice & Probiotic Drinks, Cheese Sticks for Kids, Ice Cream, Chilled Yogurt, Hot Pot Ingredient Kit, Prepared Frozen Meal, Volcanic Stone Grilled Sausage
12	Meat	Chicken, Egg Products, Poultry, Duck, Lamb, Offal, Beef, Pork & Related Products, Donkey Meat
13	Fruits & Vegetables	Strawberry, Orange, Red Dragon Fruit, Pear, Apple, Kiwifruit, Bananas, Ginger & Garlic, Kimchi & Pickled Vegetables, Fresh Fruits & Vegetables, Fresh Root Vegetables, Fresh Leafy Greens, Fresh Edible Mushrooms
14	Storage Devices	Solid State Drive (SSD), Hard Disk Drive (HDD), Memory Module (RAM), External Hard Drive Enclosure, Portable Solid State Drive, USB Flash Drive, External Hard Drive, Multi-function Card Reader, Camera Memory Card
15	Camera Accessories	Camera Battery, Photo Album, Digital Photo Frame, Instant Camera and Accessories
16	Computers	Cooler, Computer Case, Computer Hardware Accessories, Motherboard, Desktop Power Supply, Graphics Card, Processor Motherboard Combo, CPU Motherboard Combo, Assembled Desktop Computer, Monitor Stand, Monitor, Desktop Computer, Laptop Accessories, Gaming Laptop, Thin and Light Laptop, Enterprise Servers and Workstations, All-in-One Computer
17	Tablets & Smart Learning Tools	Tablet Computer, Smart LCD Writing Tablet, Electronic Dictionary Pen, Smart Translation & Learning Tools, Reading Pen, Student Tablet, E-book Reader, Smart Educational Robot
18	Networking Equipment	Wireless Network Card, Mobile Wi-Fi Router, Enterprise Network Switch, NAS Network Storage Server, Internet TV Set-Top Box, Wireless Networking Equipment, Fiber Optic Cables & Ethernet Cables, Fiber Optic Transceivers & Related Components
19	Electronic & Electrical Appliance Accessories	TV Wall Mount & Remote Control, Air Conditioner Accessories, Washing Machine Accessories, Kitchen Appliance Accessories, Refrigerator Door Seal, Tablet Protective Case & Accessories, Sockets & Converters, Mobile Phone Accessories, UPS Uninterruptible Power Supply & Cabinet, Computer Cables & Adapters, Multi-function Type-C Docking Station, HD Webcam
20	Computer & Gaming Accessories	Mechanical Keyboard & Mouse Combo, Wired and Wireless Mice, Mouse Pad, Wireless Mouse, Computer & Accessory Cleaning Tools, Game Controller, Gaming Peripherals, Retro Handheld & Arcade Consoles, Video Games
21	Audio Equipment	Gaming Headset, Senior Citizen Radio, Wireless Earbuds/Headphones, Professional Audio Equipment, Wireless Microphone, Portable Media Player, Audio Sound Systems, Language Repeater & Portable Player, Voice Recorder
22	Photography	Live Streaming & Recording Equipment, Camera Lens, DSLR Camera, Action Camera & Accessories, Mirrorless Camera, Professional HD Camcorder, Photography Equipment & Accessories, Digital Camera
23	Photography Accessories	Tripod & Gimbal, Camera & Photography Accessories, Camera Lens Accessories, DSLR Camera Bag & Dry Cabinet, Camera Lens Filter, Camera Screen Protector & Cleaning Kit, Smartphone Gimbal Stabilizer, Camera Flash & Accessories
24	Smart Devices	Electric Hoverboard, GPS Tracker/Anti-Loss Device, Smart Wearable Devices & Accessories, Smart Watch & Accessories, Smart Bracelet, Surveillance Camera, Virtual Reality Glasses, Smart Weight Scale & Body Fat Scale, Smart Watch, Drone

Figure 1: The human-annotated group structure of the dataset with $M = 24$.

Label	Group	Categories
1	Electric Kettles & Cookware	Electric Hot Plate, Automatic Herbal Medicine Pot, Wellness Kettle, Electric Kettle, Automatic Electric Tea Set, Electric Slow Cooker
2	Small Kitchen Appliances	Soy Milk Maker, Hand Mixer, Hand Mixer, Electric Meat Grinder, Panini Press, Pasta Maker, Juicer, Coffee Maker, Egg Cooker, Food Processor, High-Speed Blender, Toaster Oven, Pancake Maker
3	Large Kitchen Appliances	***Electric Heating Table***, Multi-function Electric Pot, Microwave Oven, Electric Pressure Cooker, Air Fryer, Electric Lunch Box, Food Warmer Plate, Electric Oven, Induction Cooktop, Home Food Warmer Plate, Electric Oven, Rice Cooker, Combination Steam Oven
4	Laundry and Garment Care Tools	***Washing Machine Accessories***, Washing Machine, Dryer, Washer-Dryer Combo, Steam Iron, Sanitizing Washer Dryer
5	Air Conditioning and Ventilation Systems	***Freezer***, Central Air Conditioning, Air Conditioner, Portable Air Conditioner, Air Cooler, Fresh Air System, Electric Fan
6	Cleaning Appliances	***Computer & Accessory Cleaning Tools***, Water Purifier & Air Purifier Filters, Air Purifier, Household Water Purifier, Robotic Vacuum Cleaner, Dehumidifier, Household Dust Mite Eliminator, Electric Cleaning & Steam Equipment, Vacuum Cleaner, Humidifier
7	Other Home Appliances	Wine Cooler, Refrigerator, Sterilizing Cabinet, Range Hood & Gas Stove Set, Dishwasher, Integrated Stove, Gas Stove, Solar Water Heater, Electric Water Heater, Air Source Heat Pump Water Heater, Gas Water Heater, Kitchen Appliance Accessories, Heater, Home Electronics, Water Dispenser, Tea Bar Machine, Smart Home Devices
8	Personal Care Appliances	***Fabric Shaver***, Shaver Accessories, Electric Toothbrush Replacement Heads, Hair Curling Iron/Straightener, Electric Oral Irrigator, Personal Grooming Hair Removal & Trimmer, Hair Dryer, Electric Shaver, Electric Toothbrush, Beauty Devices, Hair Cutting Tools
9	Massage Equipment	Massager, Foot Massager, Home Massage & Physical Therapy Equipment, Massage Chair, Foot Bath, Home Electric Physical Therapy Device
10	Seafood	Softshell Turtles, Canned Prepared Seafood, Other Seafood, Crabs, Seafood Related Products, Seafood Gift Boxes, Seaweed Products, Shrimp & Prawns, Dried Sea Cucumbers, Seafood, Dried Seafood
11	Refrigerated & Frozen Foods	Pizza, Ready-to-Cook Baked Goods, Dumplings, Portuguese Egg Tarts, Birthday Cakes, Frozen Foods, Juice & Probiotic Drinks, Cheese Sticks for Kids, Ice Cream, Chilled Yogurt, Prepared Frozen Meal
12	Meat	***Hot Pot Ingredient Kit***, ***Volcanic Stone Grilled Sausage***, Chicken, Egg Products, Poultry, Duck, Lamb, Offal, Beef, Pork & Related Products, Donkey Meat
13	Fruits & Vegetables	Strawberry, Orange, Red Dragon Fruit, Pear, Apple, Kiwifruit, Bananas, Ginger & Garlic, Kimchi & Pickled Vegetables, Fresh Fruits & Vegetables, Fresh Root Vegetables, Fresh Leafy Greens, Fresh Edible Mushrooms
14	Storage Devices	***NAS Network Storage Server***, Solid State Drive (SSD), Hard Disk Drive (HDD), Memory Module (RAM), External Hard Drive Enclosure, Portable Solid State Drive, USB Flash Drive, External Hard Drive, Multi-function Card Reader, Camera Memory Card
15	Camera Accessories	Camera Battery, Photo Album, Digital Photo Frame, Instant Camera and Accessories
16	Computers	Motherboard, Graphics Card, Processor Motherboard Combo, CPU Motherboard Combo, Assembled Desktop Computer, Desktop Computer, Gaming Laptop, Thin and Light Laptop, Enterprise Servers and Workstations, All-in-One Computer
17	Tablets & Smart Learning Tools	Smart LCD Writing Tablet, Electronic Dictionary Pen, Smart Translation & Learning Tools, Reading Pen, Student Tablet, E-book Reader, Smart Educational Robot
18	Networking Equipment	***Computer Cables & Adapters***, ***Multi-function Type-C Docking Station***, Wireless Network Card, Mobile Wi-Fi Router, Enterprise Network Switch, Wireless Networking Equipment, Fiber Optic Cables & Ethernet Cables, Fiber Optic Transceivers & Related Components
19	Electronic & Electrical Appliance Accessories	***Cooler***, ***Computer Case***, ***Computer Hardware Accessories***, ***Desktop Power Supply***, ***Monitor Stand***, TV Wall Mount & Remote Control, Air Conditioner Accessories, Kitchen Appliance Accessories, Refrigerator Door Seal, Sockets & Converters, UPS Uninterruptible Power Supply & Cabinet
20	Computer & Gaming Accessories	***Laptop Accessories***, ***Tablet Computer***, ***Tablet Protective Case & Accessories***, Mechanical Keyboard & Mouse Combo, Wired and Wireless Mice, Mouse Pad, Wireless Mouse, Game Controller, Gaming Peripherals, Retro Handheld & Arcade Consoles, Video Games
21	Audio Equipment	***Desktop Telephone***, Gaming Headset, Senior Citizen Radio, Wireless Earbuds/Headphones, Professional Audio Equipment, Wireless Microphone, Portable Media Player, Audio Sound Systems, Language Repeater & Portable Player, Voice Recorder
22	Photography	***Smart TV***, ***Monitor***, ***Internet TV Set-Top Box***, ***HD Webcam***, ***Surveillance Camera***, ***Smartphone Gimbal Stabilizer***, ***Virtual Reality Glasses***, ***Drone***, Live Streaming & Recording Equipment, Action Camera & Accessories, Professional HD Camcorder, Photography Equipment & Accessories
23	Photography Accessories	***Camera Lens***, ***DSLR Camera***, ***Mirrorless Camera***, ***Digital Camera***, Tripod & Gimbal, Camera & Photography Accessories, Camera Lens Accessories, DSLR Camera Bag & Dry Cabinet, Camera Lens Filter, Camera Screen Protector & Cleaning Kit, Camera Flash & Accessories
24	Smart Devices	***Electronic Scale***, ***Mobile Phone Accessories***, Electric Hoverboard, GPS Tracker/Anti-Loss Device, Smart Wearable Devices & Accessories, Smart Watch & Accessories, Smart Bracelet, Smart Weight Scale & Body Fat Scale, Smart Watch

Figure 2: The GGM-discovered group structure of the dataset with $M = 24$. Those incorrectly identified ones are highlighted with italic and bold.

REFERENCES

Boldea, O. & Magnus, J. R. (2009). Maximum likelihood estimation of the multivariate normal mixture model. *Journal of the American Statistical Association*, *104*(488), 1539–1549.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.