# Supplemental: Finite Mixtures of Multivariate Poisson-Log Normal Factor Analyzers for Clustering Count Data

Anjali Silva*
Steven J. Rothstein†
Paul D. McNicholas‡
Sanjeena Subedi§

**Mathematical Details**

A $G$-component mixture of MPLN factor analyzers has the distribution

$$\begin{aligned}
f(\mathbf{y};\boldsymbol{\Theta}) &= \sum_{g=1}^{G} \pi_g f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g, \boldsymbol{\Psi}_g) \\
&= \sum_{g=1}^{G} \pi_g \int_{\mathbb{R}^p} \left( \prod_{j=1}^{p} f(y_{ij}|\theta_{ijg}, s_j) \right) f(\boldsymbol{\theta}_{ig}|\boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g, \boldsymbol{\Psi}_g)\ d\boldsymbol{\theta}_{ig}.
\end{aligned}$$

Here $\boldsymbol{\Theta} = (\pi_1, \ldots, \pi_G, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_G, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_G, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_G)$. At the first stage of the MCMC-AECM algorithm, when estimating $\boldsymbol{\vartheta}_1 = (\boldsymbol{\theta}_g, \pi_g, \boldsymbol{\mu}_g; g = 1, \ldots, G)$, the group labels $z_{ig}$ and $\boldsymbol{\theta}_{ig}, i = 1, \ldots, n; g = 1, \ldots, G$ are the missing data. So the complete data likelihood for the mixture model is

$$L_{c1}(\boldsymbol{\vartheta}_1) = \prod_{i=1}^{n} \prod_{g=1}^{G} \left[ \pi_g \left( \prod_{j=1}^{p} f(y_{ij}|\theta_{ijg}, s_j) \right) f(\boldsymbol{\theta}_{ig}|\boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g, \boldsymbol{\Psi}_g)\ d\boldsymbol{\theta}_{ig} \right]^{z_{ig}}.$$

*Department of Mathematics and Statistics, University of Guelph, Guelph, ON, Canada.
†Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada.
‡Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada.
§Department of Mathematical Sciences, Binghamton University, Binghamton, New York, USA.

The complete data log-likelihood for the mixture model is

$$
\begin{aligned}
l_{c1}(\boldsymbol{\vartheta}_1) &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big[\log\pi_g + \sum_{j=1}^{p}\log(f(y_{ij}|\theta_{ijg}, s_j)) + \log(f(\boldsymbol{\theta}_{ig}|\boldsymbol{\mu}_g, \boldsymbol{\Lambda}_g, \boldsymbol{\Psi}_g))\Big],\\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\log\pi_g + \sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}\Big(\log\big(\frac{\exp\{-\exp\{\theta_{ijg}+\log s_j\}\}(\exp\{\theta_{ijg}+\log s_j\})^{y_{ij}}}{y_{ij}!}\big)\Big)\\
&+ \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big(\log\big(\frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g|}}\exp\{-\frac{1}{2}[(\boldsymbol{\theta}_i-\boldsymbol{\mu}_g)(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g)^{-1}(\boldsymbol{\theta}_i-\boldsymbol{\mu}_g)']\}\big)\Big),\\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big[\log\pi_g + \sum_{j=1}^{p}\log\big(\frac{\exp\{-\exp\{\theta_{ijg}+\log s_j\}\}(\exp\{\theta_{ijg}+\log s_j\})^{y_{ij}}}{y_{ij}!}\big)\\
&+ \log\big(\frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g|}}\exp\{-\frac{1}{2}[(\boldsymbol{\theta}_i-\boldsymbol{\mu}_g)(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g)^{-1}(\boldsymbol{\theta}_i-\boldsymbol{\mu}_g)']\}\big)\Big],\\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\Big[\log\pi_g - \sum_{j=1}^{p}\exp\{\theta_{ijg}+\log s_j\} + \sum_{j=1}^{p}(\theta_{ijg}+\log s_j)y_{ij} - \sum_{j=1}^{p}\log y_{ij}!\\
&- \frac{p}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g| - \frac{1}{2}(\boldsymbol{\theta}_i-\boldsymbol{\mu}_g)(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g)^{-1}(\boldsymbol{\theta}_i-\boldsymbol{\mu}_g)'\Big],\\
&= \sum_{g=1}^{G} n_g\log\pi_g - \sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}\exp\{\theta_{ijg}+\log s_j\} + \sum_{i=1}^{n}\sum_{i=g}^{G} z_{ig}(\boldsymbol{\theta}_{ig}+\log\mathbf{s})\mathbf{y}_i'\\
&- \sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}\log y_{ij}! - \frac{np}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g\log|\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g|\\
&- \frac{1}{2}\sum_{g=1}^{G} n_g\mathrm{tr}\{\mathbf{S}_g(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'+\boldsymbol{\Psi}_g)^{-1}\},
\end{aligned}
$$

where $n_g = \sum_{i=1}^{n} z_{ig}$. Here, $\mathbf{S}_g$ represents the sample covariance matrix for component g, which has the form

$$
\mathbf{S}_g = \frac{1}{n_g}\sum_{i=1}^{n} z_{ig}\,\mathbb{E}\left((\boldsymbol{\theta}_{ig}-\boldsymbol{\mu}_g)(\boldsymbol{\theta}_{ig}-\boldsymbol{\mu}_g)'\right).
$$

At each E-step, the (conditional) expected value of $\boldsymbol{\theta}_{ig}$ and (conditional) expected value of group membership variable, $\mathbf{z}_{ig}$, are updated as

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{\theta}_{ig}|\mathbf{y}_i) &\simeq \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\theta}_{ig}^{(k)} \simeq \boldsymbol{\theta}_{ig}^{(t)},\\
\mathbb{E}(Z_{ig}|\mathbf{y}_i,\boldsymbol{\theta}_{ig},\mathbf{s}) &= \frac{\pi_g f(y_{ij}|\boldsymbol{\theta}_{ig}^{(t)},\mathbf{s})f(\boldsymbol{\theta}_{ig}|\boldsymbol{\mu}_g^{(t)},\boldsymbol{\Lambda}_g^{(t)},\boldsymbol{\Psi}_g^{(t)})}{\sum_{h=1}^{G}\pi_h^{(t)}f(y_{ij}|\boldsymbol{\theta}_{ih}^{(t)},\mathbf{s})f(\boldsymbol{\theta}_{ih}|\boldsymbol{\mu}_h^{(t)},\boldsymbol{\Lambda}_h^{(t)},\boldsymbol{\Psi}_h^{(t)})} =: z_{ig}^{(t)}.
\end{aligned} \tag{1}
$$

Note that, in the E-step, the estimates are conditioned on the current parameter estimates, hence the use of $(t)$ on the parameters in (1). Using the expected values given by (1), the expected value of the complete-data log-likelihood at first stage is

$$
\begin{aligned}
\mathcal{Q}_1 \simeq& \sum_{g=1}^{G} n_g^{(t)}\log\pi_g^{(t)} - \sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}^{(t)}\exp\{\mathbb{E}(\theta_{ijg})+\log s_j\} + \sum_{i=1}^{n}\sum_{i=g}^{G} z_{ig}^{(t)}\big(\mathbb{E}(\boldsymbol{\theta}_{ig})+\log\mathbf{s}\big)\mathbf{y}_i'\\
&-\sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}\log y_{ijg}! - \frac{np}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g^{(t)}\log|\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)\prime}+\boldsymbol{\Psi}_g^{(t)}|\\
&-\frac{1}{2}\sum_{g=1}^{G} n_g^{(t)}\mathrm{tr}\big\{\mathbf{S}_g\big(\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)\prime}+\boldsymbol{\Psi}_g^{(t)}\big)^{-1}\big\}.
\end{aligned}
$$

where $n_g^{(t)} = \sum_{i=1}^{n} z_{ig}^{(t)}$. Here, $\mathbf{S}_g$ represents the sample covariance matrix for component g, which has the form

$$
\mathbf{S}_g = \frac{1}{n_g^{(t)}}\sum_{i=1}^{n} z_{ig}^{(t)}\,\mathbb{E}\left((\boldsymbol{\theta}_{ig}-\boldsymbol{\mu}_g^{(t)})(\boldsymbol{\theta}_{ig}-\boldsymbol{\mu}_g^{(t)})'\right). \tag{2}
$$

Maximizing $\mathcal{Q}_1$ with respect to $\pi_g$ and $\boldsymbol{\mu}_g$ leads to the parameter updates,

$$
\begin{aligned}
\pi_g^{(t+1)} &= \frac{\sum_{i=1}^{n} z_{ig}^{(t)}}{n} = \frac{n_g^{(t)}}{n}\\
\boldsymbol{\mu}_g^{(t+1)} &= \frac{\sum_{i=1}^{n} z_{ig}^{(t)}\,\mathbb{E}(\boldsymbol{\theta}_{ig})}{\sum_{i=1}^{n} z_{ig}^{(t)}} = \frac{\sum_{i=1}^{n} z_{ig}^{(t)}\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\theta}_{ig}^{(k)}}{n_g^{(t)}}.
\end{aligned}
$$

At the second stage of the MCMC-AECM algorithm, when estimating $\boldsymbol{\vartheta}_2 = (\boldsymbol{\Lambda}_g,\boldsymbol{\Psi}_g; g = 1,\ldots,G)$, the group labels $\mathbf{z}_i$, the latent factors $\mathbf{u}_{ig}$ and $\boldsymbol{\theta}_{ig}, i = 1,\ldots,n; g = 1,\ldots,G$, are taken to be the missing data. Here, $\boldsymbol{\theta}_{ig}|\mathbf{u}_{ig} \sim N(\boldsymbol{\mu}_g+\boldsymbol{\Lambda}_g\mathbf{u}_{ig},\boldsymbol{\Psi}_g)$ and $\mathbf{u}_{ig} \sim \mathcal{N}(\mathbf{0},\mathbf{I}_q)$. The complete data likelihood is

$$
L_{c2}(\boldsymbol{\vartheta}_2) = \prod_{i=1}^{n}\left[\prod_{g=1}^{G}\pi_g^{(t+1)}\left(\prod_{j=1}^{p} f(y_{ij}|\theta_{ijg}^{(t)},s_j)\right) f(\boldsymbol{\theta}_{ig}|\mathbf{u}_i,\boldsymbol{\mu}_g^{(t+1)},\boldsymbol{\Lambda}_g^{(t)},\boldsymbol{\Psi}_g^{(t)})f(\mathbf{u}_i|\mathbf{0},\mathbf{I}_q)\right]^{z_{ig}^{(t)}},
$$

and the complete data log-likelihood is

$$
\begin{aligned}
l_{c2}(\boldsymbol{\vartheta}_2) &= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}^{(t)} \log\left[\pi_g^{(t+1)}\left(\prod_{j=1}^{p} f(y_{ij}|\theta_{ijg}^{(t)}, s_j)\right) f(\boldsymbol{\theta}_{ig}|\mathbf{u}_i, \boldsymbol{\mu}_g^{(t+1)}, \boldsymbol{\Lambda}_g^{(t)}, \boldsymbol{\Psi}_g^{(t)}) f(\mathbf{u}_i|\mathbf{0}, \mathbf{I}_q)\right], \\
&= \sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}^{(t)} \Bigg[\log \pi_g^{(t+1)} + \left(\sum_{j=1}^{p} \log f(y_{ij}|\theta_{ijg}^{(t)}, s_j)\right) + \log f(\boldsymbol{\theta}_{ig}|\mathbf{u}_i, \boldsymbol{\mu}_g^{(t+1)}, \boldsymbol{\Lambda}_g^{(t)}, \boldsymbol{\Psi}_g^{(t)}) \\
&+ \log f(\mathbf{u}_i|\mathbf{0}, \mathbf{I}_q)\Bigg], \\
&= \sum_{g=1}^{G}\Bigg[n_g^{(t)} \log \pi_g^{(t+1)} - \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ig}^{(t)} \exp\{\theta_{ijg}^{(t)} + \log s_j\} + \sum_{i=1}^{n} z_{ig}(\boldsymbol{\theta}_{ig}^{(t)} + \log \mathbf{s})\mathbf{y}_i' \\
&- \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ig}^{(t)} \log y_{ij}! - \frac{n_g}{2}\log|\boldsymbol{\Psi}_g^{(t)}| - \frac{n_g}{2}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)} - \boldsymbol{\Lambda}_g^{(t)}\mathbf{u}_i)'\boldsymbol{\Psi}_g^{(t)^{-1}}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)} - \boldsymbol{\Lambda}_g^{(t)}\mathbf{u}_i)\Bigg], \\
&= \sum_{g=1}^{G}\Bigg[n_g^{(t)} \log \pi_g^{(t+1)} - \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ig}^{(t)} \exp\{\theta_{ijg}^{(t)} + \log s_j\} + \sum_{i=1}^{n} z_{ig}^{(t)}(\boldsymbol{\theta}_{ig}^{(t)} + \log \mathbf{s})\mathbf{y}_i' \\
&- \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ig}^{(t)} \log y_{ij}! - \frac{n_g^{(t)}}{2}\log|\boldsymbol{\Psi}_g^{(t)}| - \frac{n_g^{(t)}}{2}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}^{(t+1)})'\boldsymbol{\Psi}_g^{(t)^{-1}}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)}) \\
&+ \sum_{i=1}^{n} z_{ig}^{(t)}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})'\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\mathbf{u}_i - \frac{1}{2}\sum_{i=1}^{n}\mathbf{u}_i'\boldsymbol{\Lambda}_g^{(t)\prime}\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\mathbf{u}_i\Bigg], \\
&= \sum_{g=1}^{G}\Bigg[n_g^{(t)} \log \pi_g^{(t+1)} - \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ig}^{(t)} \exp\{\theta_{ijg}^{(t)} + \log s_j\} + \sum_{i=1}^{n} z_{ig}^{(t)}(\boldsymbol{\theta}_g^{(t)} + \log \mathbf{s})\mathbf{y}_i' \\
&- \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ig}^{(t)} \log y_{ij}! - \frac{n_g^{(t)}}{2}\log|\boldsymbol{\Psi}_g^{(t)}| - \frac{n_g^{(t)}}{2}\mathrm{tr}\{\boldsymbol{\Psi}_g^{(t)^{-1}}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})'\} \\
&+ \sum_{i=1}^{n} z_{ig}^{(t)}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})'\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\mathbf{u}_i - \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Lambda}_g^{(t)\prime}\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\sum_{i=1}^{n}\mathbf{u}_i\mathbf{u}_i'\}\Bigg], \\
&= C + \sum_{g=1}^{G}\Bigg[-\frac{n_g^{(t)}}{2}\log|\boldsymbol{\Psi}_g^{(t)}| - \frac{n_g^{(t)}}{2}\mathrm{tr}\{\boldsymbol{\Psi}_g^{(t)^{-1}}\mathbf{S}_g\} + \sum_{i=1}^{n} z_{ig}^{(t)}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})'\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\mathbf{u}_i \\
&- \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Lambda}_g^{(t)\prime}\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\sum_{i=1}^{n} z_{ig}^{(t)}\mathbf{u}_i\mathbf{u}_i'\Big\}\Bigg],
\end{aligned}
$$

where C is a constant with respect to $\boldsymbol{\Lambda}_g$ and $\boldsymbol{\Psi}_g$. The expected value of the complete-data log-likelihood at second stage is

$$\begin{aligned}\mathcal{Q}_2 \simeq C + \sum_{g=1}^{G}\Big( &- \frac{n_g^{(t)}}{2}\log|\boldsymbol{\Psi}_g^{(t)}| - \frac{n_g^{(t)}}{2}\mathrm{tr}\{\boldsymbol{\Psi}_g^{(t)^{-1}}\mathbf{S}_g\} \\ &+ \sum_{i=1}^{n} z_{ig}^{(t)}(\boldsymbol{\theta}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})'\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\,\mathbb{E}(\mathbf{u}_i|\boldsymbol{\theta}_i^{(t)}, z_{ig}^{(t)} = 1) \\ &- \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Lambda}_g^{(t)\prime}\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\sum_{i=1}^{n} z_{ig}^{(t)}\,\mathbb{E}(\mathbf{u}_i\mathbf{u}_i'|\boldsymbol{\theta}_i^{(t)}, z_{ig}^{(t)} = 1)\Big\}\Big).\end{aligned} \tag{3}$$

Consider the joint distribution of $\boldsymbol{\theta}_{ig}$ and $\mathbf{u}_{ig}$,

$$\begin{bmatrix} \boldsymbol{\theta}_{ig} \\ \mathbf{U}_{ig} \end{bmatrix} \sim N\left( \begin{bmatrix} \boldsymbol{\mu}_g \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g & \boldsymbol{\Lambda}_g \\ \boldsymbol{\Lambda}_g' & \mathbf{I}_q \end{bmatrix} \right).$$

Given this information, the following can be calculated as

$$\mathbb{E}(\mathbf{U}_{ig}|\boldsymbol{\theta}_{ig}) = \boldsymbol{\mu}_{u_{ig}u_{ig}} + \boldsymbol{\Sigma}_{u_{ig}\theta_{ig}}\boldsymbol{\Sigma}_{\theta_{ig}\theta_{ig}}^{-1}(\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_{\theta_{ig}}),$$

and

$$\mathbb{V}\mathrm{ar}(\mathbf{U}_{ig}|\boldsymbol{\theta}_{ig}) = \boldsymbol{\Sigma}_{u_{ig}u_{ig}} - \boldsymbol{\Sigma}_{u_{ig}\theta_{ig}}\boldsymbol{\Sigma}_{\theta_{ig}\theta_{ig}}^{-1}\boldsymbol{\Sigma}_{\theta_{ig}u_{ig}}.$$

Accordingly, this results

$$\begin{aligned}&\mathbb{E}(\mathbf{U}_{ig}|\boldsymbol{\theta}_{ig}) = \boldsymbol{\Lambda}_g'\big(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g\big)^{-1}\big(\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g\big) \\ &\mathbb{E}(\mathbf{U}_{ig}\mathbf{U}_{ig}'|\boldsymbol{\theta}_{ig}, z_i = 1) = \mathbb{V}\mathrm{ar}(\mathbf{U}_{ig}|\boldsymbol{\theta}_{ig}) + \mathbb{E}(\mathbf{U}_{ig}|\boldsymbol{\theta}_{ig})\,\mathbb{E}(\mathbf{U}_{ig}|\boldsymbol{\theta}_{ig})' \\ &\qquad = \big(\mathbf{I}_q + \boldsymbol{\Lambda}_g'\boldsymbol{\Psi}_g^{-1}\boldsymbol{\Lambda}_g\big)^{-1} + \Big(\boldsymbol{\Lambda}_g'\big(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g\big)^{-1}\big(\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g\big)\Big)\Big(\boldsymbol{\Lambda}_g'\big(\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g\big)^{-1}\big(\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g\big)\Big)'\end{aligned}$$

Therefore, (3) can be written as

$$\begin{aligned}\mathcal{Q}_2 &\simeq C + \frac{1}{2}\sum_{g=1}^{G} n_g^{(t)}\Big[\log|\boldsymbol{\Psi}_g^{(t)^{-1}}| - \mathrm{tr}\big\{\boldsymbol{\Psi}_g^{(t)^{-1}}\mathbf{S}_g\big\} + 2\mathrm{tr}\big\{\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\big\} - \mathrm{tr}\big\{\boldsymbol{\Lambda}_g^{(t)\prime}\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Phi}_g^{(t)}\big\}\Big], \\ &\simeq C + \frac{n}{2}\Big[\log|\boldsymbol{\Psi}_g^{(t)^{-1}}| - \mathrm{tr}\big\{\boldsymbol{\Psi}_g^{(t)^{-1}}\mathbf{S}_g\big\} + 2\mathrm{tr}\big\{\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\big\} - \mathrm{tr}\big\{\boldsymbol{\Lambda}_g^{(t)\prime}\boldsymbol{\Psi}_g^{(t)^{-1}}\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Phi}_g^{(t)}\big\}\Big],\end{aligned} \tag{4}$$

where $\sum_{g=1}^{G} n_g^{(t)} = n$, $\hat{\boldsymbol{\beta}}_g$ is a $q \times p$ matrix is given by

$$\hat{\boldsymbol{\beta}}_g^{(t)} = \boldsymbol{\Lambda}_g^{(t)\prime}\big(\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)\prime} + \boldsymbol{\Psi}_g^{(t)}\big)^{-1},$$

and $\mathbf{\Phi}_g$ is a symmetric $q \times q$ matrix given by

$$\mathbf{\Phi}_g^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{\Lambda}_g^{(t)} + \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime}.$$

Note, the $\boldsymbol{\mu}_g^{(t+1)}$ replaces $\boldsymbol{\mu}_g^{(t)}$ in $\mathbf{S}_g$, cf. (2). Differentiating (4) with respect to $\mathbf{\Lambda}_g$ and $\mathbf{\Psi}_g^{-1}$, respectively, leads to

$$S_1(\mathbf{\Lambda}_g, \mathbf{\Psi}_g) = \frac{\partial \mathcal{Q}_2(\mathbf{\Lambda}_g, \mathbf{\Psi}_g)}{\partial \mathbf{\Lambda}_g} = \sum_{g=1}^{G} n_g^{(t)} \Big[ \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime} - \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{\Lambda}_g^{(t)} \mathbf{\Phi}_g^{(t)} \Big],$$

$$S_2(\mathbf{\Lambda}_g, \mathbf{\Psi}_g) = \frac{\partial \mathcal{Q}_2(\mathbf{\Lambda}_g, \mathbf{\Psi}_g)}{\partial \mathbf{\Psi}_g^{-1}} = \frac{n_g^{(t)}}{2} \Big[ \mathbf{\Psi}_g^{(t)} - \mathbf{S}_g' + 2\mathbf{\Lambda}_g^{(t)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g - \mathbf{\Lambda}_g^{(t)} \mathbf{\Phi'}_g^{(t)} \mathbf{\Lambda}_g^{(t)\prime} \Big].$$

Solving $S_1(\mathbf{\Lambda}_g^{(t+1)}, \mathbf{\Psi}_g^{(t+1)}) = 0$ and $\operatorname{diag}\{S_2(\mathbf{\Lambda}_g^{(t+1)}, \mathbf{\Psi}_g^{(t+1)})\} = 0$, leads to

$$\mathbf{\Lambda}_g^{(t+1)} = \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime} \mathbf{\Phi}_g^{(t)^{-1}},$$

and

$$\mathbf{\Psi}_g^{(t+1)} = \operatorname{diag}\{\mathbf{S}_g - \mathbf{\Lambda}_g^{(t+1)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g\}.$$

The form of complete-data log-likelihood and the parameter estimates will vary depending on which of the four models in the PMPLNFA family is under consideration.

**If equal loading matrices**: $\mathbf{\Lambda}_g = \mathbf{\Lambda}$ are assumed, then (4) can be written as

$$\mathcal{Q}_2 \simeq C + \frac{1}{2} \sum_{g=1}^{G} n_g^{(t)} \Big[ \log |\mathbf{\Psi}_g^{(t)^{-1}}| - \operatorname{tr}\big\{ \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{S}_g \big\} + 2\operatorname{tr}\big\{ \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{\Lambda}^{(t)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g \big\} - \operatorname{tr}\big\{ \mathbf{\Lambda}^{(t)\prime} \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{\Lambda}^{(t)} \mathbf{\Phi}_g^{(t)} \big\} \Big], \tag{5}$$

where $\hat{\boldsymbol{\beta}}_g^{(t)} = \mathbf{\Lambda}^{(t)\prime} \big( \mathbf{\Lambda}^{(t)} \mathbf{\Lambda}^{(t)\prime} + \mathbf{\Psi}_g^{(t)} \big)^{-1}$ and $\mathbf{\Phi}_g^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{\Lambda}^{(t)} + \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime}$. Differentiating (5) with respect to $\mathbf{\Lambda}^{(t)}$ and $\mathbf{\Psi}_g^{(t)^{-1}}$, respectively, leads to

$$S_3(\mathbf{\Lambda}, \mathbf{\Psi}_g) = \frac{\partial \mathcal{Q}_2(\mathbf{\Lambda}, \mathbf{\Psi}_g)}{\partial \mathbf{\Lambda}} = \sum_{g=1}^{G} n_g^{(t)} \Big[ \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime} - \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{\Lambda}^{(t)} \mathbf{\Phi}_g^{(t)} \Big],$$

$$S_4(\mathbf{\Lambda}, \mathbf{\Psi}_g) = \frac{\partial \mathcal{Q}_2(\mathbf{\Lambda}, \mathbf{\Psi}_g)}{\partial \mathbf{\Psi}_g^{-1}} = \frac{n_g^{(t)}}{2} \Big[ \mathbf{\Psi}_g^{(t)} - \mathbf{S}_g' + 2\mathbf{\Lambda}^{(t)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g - \mathbf{\Lambda}^{(t)} \mathbf{\Phi}_g^{(t)\prime} \mathbf{\Lambda}^{(t)\prime} \Big].$$

Setting $S_3(\mathbf{\Lambda}^{(t+1)}, \mathbf{\Psi}_g^{(t+1)}) = 0$ leads to

$$\sum_{g=1}^{G} n_g \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{\Lambda}^{(t+1)} \mathbf{\Phi}_g^{(t)} = \sum_{g=1}^{G} n_g^{(t)} \mathbf{\Psi}_g^{(t)^{-1}} \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime}, \tag{6}$$

which must be solved for $\mathbf{\Lambda}^{(t+1)}$ in a row-by-row manner. This slows the fitting of this model. Let $\lambda_i^{(t)} = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i1})$ represent the $i$th row of the matrix $\mathbf{\Lambda}$ and let $r_i$ represent the $i$th row of the matrix on the right-hand side of (6). Here $i$th row of (6) can be written as

$$\lambda_i^{(t+1)} \sum_{g=1}^{G} \frac{n_g^{(t)}}{\psi_{g(i)}^{(t)}} \mathbf{\Phi}_g^{(t)} = \mathbf{r}_i,$$

where $\psi_{g(i)}^{(t)}$ is the $i$th entry along the diagonal of $\mathbf{\Psi}_g^{(t)}$. Hence,

$$\lambda_i^{(t+1)} = \mathbf{r}_i \Big( \sum_{g=1}^{G} \frac{n_g^{(t)}}{\psi_{g(i)}^{(t)}} \mathbf{\Phi}_g^{(t)} \Big)^{-1},$$

for $i = 1, \ldots, p$.

Setting $\text{diag}\{S_4(\mathbf{\Lambda}^{(t+1)}, \mathbf{\Psi}_g^{(t+1)})\} = 0$ and solving leads to

$$\mathbf{\Psi}_g^{(t+1)} = \text{diag}\{\mathbf{S}_g - 2\mathbf{\Lambda}^{(t+1)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g + \mathbf{\Lambda}^{(t+1)} \mathbf{\Phi}_g^{(t)} (\mathbf{\Lambda}^{(t+1)})'\}.$$

**If equal error variance**: $\mathbf{\Psi}_g = \mathbf{\Psi}$ are assumed, then (4) can be written as

$$\mathcal{Q}_2 \simeq C + \frac{1}{2} \sum_{g=1}^{G} n_g^{(t)} \Big[ \log |\mathbf{\Psi}^{(t)^{-1}}| - \text{tr}\big\{\mathbf{\Psi}^{(t)^{-1}} \mathbf{S}_g\big\} + 2\text{tr}\big\{\mathbf{\Psi}^{(t)^{-1}} \mathbf{\Lambda}_g^{(t)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g\big\} - \text{tr}\big\{\mathbf{\Lambda}_g^{(t)\prime} \mathbf{\Psi}^{(t)^{-1}} \mathbf{\Lambda}_g^{(t)} \mathbf{\Phi}_g^{(t)}\big\}\Big], \tag{7}$$

where $\hat{\boldsymbol{\beta}}_g^{(t)} = \mathbf{\Lambda}_g^{(t)\prime} \big(\mathbf{\Lambda}_g^{(t)} \mathbf{\Lambda}_g^{(t)\prime} + \mathbf{\Psi}^{(t)}\big)^{-1}$ and $\mathbf{\Phi}_g^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{\Lambda}_g^{(t)} + \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime}$. Differentiating (7) with respect to $\mathbf{\Lambda}_g^{(t)}$ and $\mathbf{\Psi}^{(t)^{-1}}$, respectively, leads to

$$S_5(\mathbf{\Lambda}_g, \mathbf{\Psi}) = \frac{\partial \mathcal{Q}_2(\mathbf{\Lambda}_g, \mathbf{\Psi})}{\partial \mathbf{\Lambda}_g} = \sum_{g=1}^{G} n_g^{(t)} \mathbf{\Psi}^{(t)^{-1}} \Big[\mathbf{S}_g \hat{\beta}_g^{(t)\prime} - \mathbf{\Lambda}_g^{(t)} \mathbf{\Phi}_g^{(t)}\Big],$$

$$S_6(\mathbf{\Lambda}_g, \mathbf{\Psi}) = \frac{\partial \mathcal{Q}_2(\mathbf{\Lambda}_g, \mathbf{\Psi})}{\partial \mathbf{\Psi}^{-1}} = \frac{n}{2} \mathbf{\Psi}^{(t)} - \frac{1}{2} \sum_{g=1}^{G} n_g^{(t)} \Big[\mathbf{S}_g' - 2\mathbf{\Lambda}_g^{(t)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g + \mathbf{\Lambda}_g^{(t)} \mathbf{\Phi}_g^{(t)\prime} \mathbf{\Lambda}_g^{(t)\prime}\Big].$$

Setting $S_5(\mathbf{\Lambda}_g^{(t+1)}, \mathbf{\Psi}^{(t+1)}) = 0$ leads to

$$\mathbf{\Lambda}_g^{(t+1)} = \mathbf{S}_g \hat{\boldsymbol{\beta}}_g^{(t)\prime} \mathbf{\Phi}_g^{(t)^{-1}},$$

and setting $S_6(\mathbf{\Lambda}_g^{(t+1)}, \mathbf{\Psi}^{(t+1)}) = 0$ leads to

$$\mathbf{\Psi}^{(t+1)} = \sum_{g=1}^{G} \pi_g^{(t)} \text{diag}\{\mathbf{S}_g - \mathbf{\Lambda}_g^{(t+1)} \hat{\boldsymbol{\beta}}_g^{(t)} \mathbf{S}_g\}.$$

By combining these constraints, a set of four models can be obtained as described in Table 4.1, where "constrained" means $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}$ in the loading matrix term and $\boldsymbol{\Psi}_g = \boldsymbol{\Psi}$ in the error variance term.

Estimation procedure for four covariance structures:
**Model UU**: no constraint is assumed.

$$\hat{\boldsymbol{\beta}}_g^{(t)} = \boldsymbol{\Lambda}_g^{(t)\prime}\big(\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)\prime} + \boldsymbol{\Psi}_g^{(t)}\big)^{-1},$$

$$\boldsymbol{\Phi}_g^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)} + \hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\hat{\boldsymbol{\beta}}_g^{(t)\prime},$$

$$\boldsymbol{\Lambda}_g^{(t+1)} = \mathbf{S}_g\hat{\boldsymbol{\beta}}_g^{(t)\prime}\boldsymbol{\Phi}_g^{(t)^{-1}},$$

$$\boldsymbol{\Psi}_g^{(t+1)} = \mathrm{diag}\{\mathbf{S}_g - \boldsymbol{\Lambda}_g^{(t+1)}\hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\}.$$

**Model CU**: Assume $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}$.

$$\hat{\boldsymbol{\beta}}_g^{(t)} = \boldsymbol{\Lambda}^{(t)\prime}\big(\boldsymbol{\Lambda}^{(t)}\boldsymbol{\Lambda}^{(t)\prime} + \boldsymbol{\Psi}_g^{(t)}\big)^{-1},$$

$$\boldsymbol{\Phi}_g^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}_g^{(t)}\boldsymbol{\Lambda}^{(t)} + \hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\hat{\boldsymbol{\beta}}_g^{(t)\prime},$$

$$\lambda_i^{(t+1)} = \mathbf{r}_i\Big(\sum_{g=1}^{G}\frac{n_g^{(t)}}{\psi_{g(i)}^{(t)}}\boldsymbol{\Phi}_g^{(t)}\Big)^{-1},$$

$$\boldsymbol{\Psi}_g^{(t+1)} = \mathrm{diag}\{\mathbf{S}_g - 2\boldsymbol{\Lambda}^{(t+1)}\hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g + \boldsymbol{\Lambda}^{(t+1)}\boldsymbol{\Phi}_g^{(t)}(\boldsymbol{\Lambda}^{(t+1)})'\}.$$

**Model UC**: Assume $\boldsymbol{\Psi}_g = \boldsymbol{\Psi}$.

$$\hat{\boldsymbol{\beta}}_g^{(t)} = \boldsymbol{\Lambda}_g^{(t)\prime}\big(\boldsymbol{\Lambda}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)\prime} + \boldsymbol{\Psi}^{(t)}\big)^{-1},$$

$$\boldsymbol{\Phi}_g^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}_g^{(t)}\boldsymbol{\Lambda}_g^{(t)} + \hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\hat{\boldsymbol{\beta}}_g^{(t)\prime},$$

$$\boldsymbol{\Lambda}_g^{(t+1)} = \mathbf{S}_g\hat{\boldsymbol{\beta}}_g^{(t)\prime}\boldsymbol{\Phi}_g^{-1},$$

$$\boldsymbol{\Psi}^{(t+1)} = \sum_{g=1}^{G}\pi_g^{(t)}\mathrm{diag}\{\mathbf{S}_g - \boldsymbol{\Lambda}_g^{(t+1)}\hat{\boldsymbol{\beta}}_g^{(t)}\mathbf{S}_g\}.$$

**Model CC**: Assume $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}_g = \boldsymbol{\Psi}$. Here $\tilde{\mathbf{S}} = \sum_{g=1}^{G}\pi_g^{(t)}\mathbf{S}_g$.

$$\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{\Lambda}^{(t)\prime}\big(\boldsymbol{\Lambda}^{(t)}\boldsymbol{\Lambda}^{(t)\prime} + \boldsymbol{\Psi}^{(t)}\big)^{-1},$$

$$\tilde{\boldsymbol{\Phi}}^{(t)} = \mathbf{I}_q - \hat{\boldsymbol{\beta}}^{(t)}\boldsymbol{\Lambda}^{(t)} + \hat{\boldsymbol{\beta}}^{(t)}\tilde{\mathbf{S}}\hat{\boldsymbol{\beta}}^{(t)\prime},$$

$$\boldsymbol{\Lambda}^{(t+1)} = \tilde{\mathbf{S}}\hat{\boldsymbol{\beta}}^{(t)\prime}\tilde{\boldsymbol{\Phi}}^{-1},$$

$$\boldsymbol{\Psi}^{(t+1)} = \mathrm{diag}\{\tilde{\mathbf{S}} - \boldsymbol{\Lambda}^{(t+1)}\hat{\boldsymbol{\beta}}^{(t)}\tilde{\mathbf{S}}\}.$$