I. More Inference details on NMoG-LRMF

The full likelihood of the proposed NMoG-LRMF model is expressed as:

$$\begin{aligned}
&p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\\
=&p(\boldsymbol{Y}|\boldsymbol{\mathcal{Z}},\boldsymbol{U},\boldsymbol{V},\boldsymbol{Y},\boldsymbol{\mu},\boldsymbol{\tau})p(\boldsymbol{\mathcal{Z}}|\boldsymbol{\pi})p(\boldsymbol{U}|\boldsymbol{\gamma})\\
&p(\boldsymbol{V}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\boldsymbol{\mu},\boldsymbol{\tau}|\boldsymbol{d})p(\boldsymbol{\pi})p(\boldsymbol{d})\\
=&\prod\nolimits_{i,j,k}\mathcal{N}(y_{ij}|\mu_{jk}+\boldsymbol{u}_{i\cdot}\boldsymbol{v}_{j\cdot}^T,\tau_{jk}^{-1})^{z_{ijk}}\prod\nolimits_{i,j}\mathrm{Mult}(\boldsymbol{z}_{ij}|\boldsymbol{\pi}_j)\\
&\prod\nolimits_{l}\{\mathcal{N}(\boldsymbol{u}_{\cdot l}|0,\gamma_l^{-1}I_N)\mathcal{N}(\boldsymbol{v}_{\cdot l}|0,\gamma_l^{-1}I_B)\mathrm{Gam}(\gamma_l|\xi_0,\delta_0)\}\\
&\prod\nolimits_{j,k}\{\mathcal{N}(\mu_{jk}|m_0,(\beta_0\tau_{jk})^{-1})\mathrm{Gam}(\tau_{jk}|c_0,d_{jk})\}\\
&\prod\nolimits_{j}\mathrm{Dir}(\boldsymbol{\pi}_j|\alpha_0)\prod\nolimits_{j}\mathrm{Gam}(d_{jk}|\eta_0,\lambda_0).
\end{aligned}$$

In the maintext, we have introduced the variational inference to calculate posterior of this model and supposed the approximation of posterior have a factorized form as follow:

$$\begin{aligned}
q(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d})=\prod\nolimits_{i}q(\boldsymbol{u}_{i\cdot})\prod\nolimits_{j}q(\boldsymbol{v}_{j\cdot})\\
\prod\nolimits_{ij}q(\boldsymbol{z}_{ij})\prod\nolimits_{j}q(\boldsymbol{\mu_j},\boldsymbol{\tau_j})q(\boldsymbol{\pi_j})\prod\nolimits_{l}q(\gamma_l)q(d),
\end{aligned}$$

Next, we give detailed deduction of each factorized distributions involved in posterior. $\langle f(\Theta)\rangle_{\setminus\Theta_i}$ denotes the expectation of $f(\Theta)$ on set of Θ with Θ_i removed.

Infer $\boldsymbol{\mu}_j,\boldsymbol{\tau}_j$:

$$\begin{aligned}
&\ln q^*(\boldsymbol{\mu}_j,\boldsymbol{\tau}_j)\\
=&\langle\ln p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\rangle_{\setminus\boldsymbol{\mu},\boldsymbol{\tau}}+\mathrm{const}\\
=&\sum\nolimits_{i,k}-\frac{1}{2}\langle z_{ijk}\rangle\langle\tau_{jk}(\mu_{jk}-Y_{ij}-u_{i\cdot}v_{j\cdot}^T)^2\rangle-\\
&\sum\nolimits_{i,k}\frac{1}{2}\beta_0\tau_{jk}(\mu_{jk}-m_0)^2+(\frac{1}{2}\langle z_{ijk}\rangle+c_0-\frac{1}{2})\ln\tau_{jk}\\
&-\sum\nolimits_{k}\langle d\rangle\tau_{jk}+\mathrm{const}\\
=&\sum\nolimits_{k}\frac{1}{2}\ln(\beta_{jk}\tau_{jk})-\frac{\beta_{jk}\tau_{jk}}{2}(\mu_{jk}-m_{jk})^2\\
&+\sum\nolimits_{k}(c_{jk}-1)\tau_{jk}-d_{jk}\tau_{jk}+\mathrm{const},
\end{aligned}\tag{1}$$

where,

$$m_{jk}=\frac{1}{\beta_{jk}}\{\beta_0m_0\sum\nolimits_{i}\langle z_{ijk}\rangle(Y_{ij}-\langle\mu_{jk}\rangle)\},$$

$$\beta_{jk}=\beta_0+\sum\nolimits_{i}\langle z_{ijk}\rangle,\ c_{jk}=c_0+\frac{1}{2}\sum\nolimits_{i}\langle z_{ijk}\rangle,$$

$$\begin{aligned}
d_{jk}=\langle d\rangle+\frac{1}{2}\{&\sum\nolimits_{i}\langle z_{ijk}\rangle\langle(Y_{ij}-\boldsymbol{u}_{i\cdot}\boldsymbol{v}_{j\cdot}^T)^2\rangle+\beta_0m_0^2\\
&-\frac{1}{\beta_{jk}}[\sum\nolimits_{i}\langle z_{ijk}\rangle(Y_{ij}-\boldsymbol{u}_{i\cdot}\boldsymbol{v}_{j\cdot}^T)+\beta_0^2m_0^2]^2\}.
\end{aligned}$$

Taking the exponential of both sides of Eq. (1) and normalizing the right side, we obtain

$$q^*(\boldsymbol{\mu}_j,\boldsymbol{\tau}_j)=\prod\nolimits_{k}\mathcal{N}(\boldsymbol{\mu}_{jk}|m_{jk},\frac{1}{\beta_{jk}\tau_{jk}})\mathrm{Gam}(\tau_{jk}|c_{jk},d_{jk}).$$

Infer $\boldsymbol{z}_{ij}$,

$$\begin{aligned}
&\ln q^*(\boldsymbol{z}_{ij})\\
=&\langle\ln p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\rangle_{\setminus\boldsymbol{\mathcal{Z}}}+\mathrm{const}\\
=&\frac{1}{2}\sum_{k}\boldsymbol{z}_{ijk}[2\langle\ln\pi_{jk}\rangle-\ln(2\pi)+\langle\ln\tau_{jk}\rangle\\
&-\langle\tau_{jk}(Y_{ij}-\boldsymbol{u}_{i\cdot}\boldsymbol{v}_{j\cdot}^T-\mu_{jk})^2\rangle]+\mathrm{const},
\end{aligned}\tag{2}$$

for conveniently description, we let

$$2\ln\rho_{ijk}=2\langle\ln\pi_{jk}\rangle-\ln(2\pi)+\langle\ln\tau_{jk}\rangle-\langle\tau_{jk}(Y_{ij}-\boldsymbol{u}_{i\cdot}\boldsymbol{v}_{j\cdot}^T-\mu_{jk})^2\rangle,$$

then we take the exponential of both sides of Eq. (2), and normalizing it, we have the following result:

$$q^*(\boldsymbol{z}_{ij})=\prod\nolimits_{k}\varrho_{ijk}^{z_{ijk}},\tag{3}$$

where $\varrho_{ijk}=\rho_{ijk}/\sum_k\rho_{ijk}$.

Infer $\boldsymbol{\pi}_j$:

$$\begin{aligned}
&\ln q^*(\boldsymbol{\pi}_j)\\
=&\langle\ln p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\rangle_{\setminus\boldsymbol{\pi}}+\mathrm{const}\\
=&\sum\nolimits_{i,k}(\langle\boldsymbol{z}_{ijk}\rangle+\alpha_{0k}-1)\ln\pi_{jk}+\mathrm{const},
\end{aligned}\tag{4}$$

similarly, we have the result:

$$q^*(\boldsymbol{\pi}_j)=\prod\nolimits_{k}\pi_{jk}^{\alpha_{jk}-1},$$

where $\alpha_{jk}=\alpha_{0k}+\sum_i\langle z_{ijk}\rangle$.

Infer d:

$$\begin{aligned}
&\ln q^*(d)\\
=&\langle\ln p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\rangle_{\setminus\boldsymbol{d}}+\mathrm{const}\\
=&(KBc_0+\eta_0-1)\ln d-(\langle\tau_{jk}\rangle+\lambda_0)d+\mathrm{const},
\end{aligned}\tag{5}$$

after taking the exponential of both sides of Eq. (5), and normalizing it, we have:

$$q^*(d)=\mathrm{Gam}(d|\eta,\lambda),$$

where $\eta=\eta_0+c_0KB$ and $\lambda=\lambda_0+\sum_{j,k}\langle\tau_{jk}\rangle$.

Infer γ_l:

$$\begin{aligned}
&\ln q^*(\gamma_l)\\
=&\langle\ln p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\rangle_{\setminus\boldsymbol{\gamma}}+\mathrm{const}\\
=&(\xi_0+\frac{m+n}{2}-1)\ln\gamma_l-(\delta_0+\frac{1}{2}\sum\nolimits_{i}\langle\boldsymbol{u_{il}^2}\rangle\\
&+\frac{1}{2}\sum\nolimits_{j}\langle\boldsymbol{v_{jl}^2}\rangle)\gamma_l+\mathrm{const},
\end{aligned}\tag{6}$$

we can easily obtain the posterior distribution of γ_l as follow:

$$q^*(\gamma_l)=\mathrm{Gam}(\gamma_r|\xi_l,\delta_l),\tag{7}$$

where

$$\begin{aligned}
&\xi_l=\xi_0+(m+n)/2,\\
&\delta_l=\delta_0+\sum\nolimits_{i}\langle u_{il}^2\rangle/2+\sum\nolimits_{j}\langle v_{jl}^2\rangle/2.
\end{aligned}$$

Infer low-rank component $\boldsymbol{U}$ and $\boldsymbol{V}$:

$$
\begin{aligned}
&\ln q^*(\boldsymbol{u_{i\cdot}}) \\
=&\langle \ln p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\mathcal{Z}},\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{\pi},\boldsymbol{\gamma},\boldsymbol{d},\boldsymbol{Y})\rangle_{\setminus \boldsymbol{u_{i\cdot}}} + \text{const} \\
=&\sum\nolimits_{j,k} -\frac{1}{2}\langle z_{ijk}\rangle\langle \tau_{jk}(Y_{ij}-\mu_{jk}-\boldsymbol{u_{i\cdot}}\boldsymbol{v_{j\cdot}^T})^2\rangle - \frac{1}{2}\boldsymbol{u_{i\cdot}}\langle\Gamma\rangle\boldsymbol{u_{i\cdot}^T} \\
&+ \text{const} \\
=&\boldsymbol{u_{i\cdot}}(\sum\nolimits_{j,k}\langle z_{ijk}\rangle\langle\tau_{jk}\rangle\langle \boldsymbol{v_{j\cdot}^T}\boldsymbol{v_{i\cdot}}\rangle + \langle\Gamma\rangle)\boldsymbol{u_{i\cdot}^T} \\
&- 2\boldsymbol{u_{i\cdot}}\sum\nolimits_{j,k}(Y_{ij}-\mu_{jk})\boldsymbol{v_{j\cdot}^T} + \text{const},
\end{aligned} \tag{8}
$$

where $\boldsymbol{\Gamma} = \text{diag}(\langle\boldsymbol{\gamma}\rangle)$. Taking the exponential of both sides of Eq. (8), and normalizing the result, we obtain the posterior distribution of $\boldsymbol{u_{i\cdot}}$:

$$
q^*(\boldsymbol{u}_{i\cdot}) = \mathcal{N}(\boldsymbol{u}_{i\cdot}|\boldsymbol{\mu}_{\boldsymbol{u}_{i\cdot}},\boldsymbol{\Sigma}_{\boldsymbol{u}_{i\cdot}}),
$$

where,

$$
\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{u}_{i\cdot}} &= \{\sum\nolimits_{j,k}\langle z_{ijk}\rangle\langle\tau_{jk}\rangle(Y_{ij}-\langle\mu_{jk}\rangle)\langle\boldsymbol{v}_{j\cdot}\rangle\}\boldsymbol{\Sigma}_{\boldsymbol{u}_{i\cdot}}, \\
\boldsymbol{\Sigma}_{\boldsymbol{u}_{i\cdot}} &= \{\sum\nolimits_{j,k}\langle z_{ijk}\rangle\langle\tau_{jk}\rangle\langle\boldsymbol{v}_{j\cdot}^T\boldsymbol{v}_{j\cdot}\rangle + \langle\boldsymbol{\Gamma}\rangle\}^{-1}.
\end{aligned}
$$

Similarly, we can obtain the posterior distribution of $\boldsymbol{v_{j\cdot}}$ as follow:

$$
q^*(\boldsymbol{v}_{j\cdot}) = \mathcal{N}(\boldsymbol{v}_{j\cdot}|\boldsymbol{\mu}_{\boldsymbol{v}_{j\cdot}},\boldsymbol{\Sigma}_{\boldsymbol{v}_{j\cdot}}),
$$

where

$$
\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{v}_{j\cdot}} &= \{\sum\nolimits_{i,k}\langle z_{ijk}\rangle\langle\tau_{jk}\rangle(Y_{ij}-\langle\mu_{jk}\rangle)\langle\boldsymbol{u}_{i\cdot}\rangle\}\boldsymbol{\Sigma}_{\boldsymbol{v}_{j\cdot}}, \\
\boldsymbol{\Sigma}_{\boldsymbol{v}_{j\cdot}} &= \{\sum\nolimits_{i,k}\langle z_{ijk}\rangle\langle\tau_{jk}\rangle\langle\boldsymbol{u}_{i\cdot}^T\boldsymbol{u}_{i\cdot}\rangle + \langle\Gamma\rangle\}^{-1}.
\end{aligned}
$$

Here, we introduce the details of calculating all involved expectations in the update equations with respect to the current variational distributions, as listed in the following:

$$
\begin{aligned}
\langle z_{ijk}\rangle =& \varrho_{ijk}, \\
\langle \mu_{jk}\rangle =& m_{jk}, \\
\langle d_{jk}\rangle =& \frac{\eta}{\lambda}, \\
\langle (y_{ij}-\boldsymbol{u}_{i\cdot}\boldsymbol{v}_{j\cdot}^T)^2\rangle =& y_{ij}^2 + \text{tr}(\langle u_{i\cdot}^T u_{i\cdot}\rangle\langle v_{j\cdot}^T v_{j\cdot}\rangle) \\
&- 2y_{ij}\langle u_{i\cdot}v_{j\cdot}^T\rangle, \\
\langle u_{i\cdot}^T u_{i\cdot}\rangle =& \boldsymbol{\Sigma}_{\boldsymbol{u}_{i\cdot}} + \langle u_{i\cdot}\rangle\langle u_{i\cdot}\rangle^T, \\
\langle v_{j\cdot}^T v_{j\cdot}\rangle =& \boldsymbol{\Sigma}_{\boldsymbol{v}_{j\cdot}} + \langle v_{j\cdot}\rangle\langle v_{j\cdot}\rangle^T, \\
\langle \ln\pi_{jk}\rangle =& \psi(\alpha_{jk}) - \psi(\hat{\alpha}), \\
\hat{\alpha} =& \sum\nolimits_{j,k}\alpha_{jk}, \\
\langle \tau_{jk}(y_{ij}-\mu_{jk}-u_{i\cdot}v_{j\cdot}^T)^2\rangle =& \langle\tau_{jk}\rangle\langle (y_{ij}-\mu_{jk}-u_{i\cdot}v_{j\cdot}^T)^2\rangle \\
\langle (y_{ij}-\mu_{jk}-u_{i\cdot}v_{j\cdot}^T)^2\rangle =& \langle (y_{ij}-u_{i\cdot}v_{j\cdot}^T)^2\rangle \\
&- \langle\mu_{jk}\rangle(y_{ij}-\langle u_{i\cdot}\rangle\langle v_{j\cdot}\rangle^T) + \langle\mu_{jk}^2\rangle, \\
\langle \mu_{jk}^2\rangle =& (\beta_{jk}\tau_{jk})^{-1} + m_{jk}^2, \\
\langle v_{j\cdot}\rangle =& \boldsymbol{\mu}_{\boldsymbol{v}_{j\cdot}}, \\
\langle v_{j\cdot}^T v_{j\cdot}\rangle =& \boldsymbol{\Sigma}_{\boldsymbol{v}_{j\cdot}} + \langle v_{j\cdot}\rangle\langle v_{j\cdot}\rangle^T, \\
\langle\Gamma\rangle =& \text{diag}(\langle\boldsymbol{\gamma}\rangle), \langle\gamma_l\rangle = \frac{\xi_l}{\delta_l}, \\
\langle u_{i\cdot}\rangle =& \boldsymbol{\mu}_{\boldsymbol{u}_{i\cdot}}, \\
\langle u_{i\cdot}^T u_{i\cdot}\rangle =& \boldsymbol{\Sigma}_{\boldsymbol{u}_{i\cdot}} + \langle u_{i\cdot}\rangle\langle u_{i\cdot}\rangle^T, \\
\langle u_{il}^2\rangle =& (\boldsymbol{\mu}_{\boldsymbol{u}_{i\cdot}})_l^2 + \sum\nolimits_s(\boldsymbol{\Sigma}_{\boldsymbol{u}_{i\cdot}})_{ls}, \\
\langle v_{jl}^2\rangle =& (\boldsymbol{\mu}_{\boldsymbol{v}_{j\cdot}})_l^2 + \sum\nolimits_s(\boldsymbol{\Sigma}_{\boldsymbol{v}_{j\cdot}})_{ls}.
\end{aligned}
$$