The deduction process of optimal variational posterior distributions with regard to parameters and hyperparameters.

*1. The variational posterior distribution of* $\{z_i\}_{i=1}^{I}$

The optimal variational distribution of $z_i$ is $q(z_i) = \text{Categorical}(\boldsymbol{\phi}_i)$, where

$$\phi_{it} \propto \exp\left(\mathbb{E}\left[\ln \beta_t + \sum_{m=1}^{t} \ln(1-\beta_m) + \frac{1}{2}\ln|\boldsymbol{\Lambda}_t| - \frac{R}{2}\ln(2\pi) - \frac{1}{2}(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t)^T \boldsymbol{\Lambda}_t (\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t)\right]\right)$$

*2. The variational posterior distribution of* $\{\beta_t\}_{t=1}^{T}$

The optimal logarithmic variational posterior distribution of $\beta_t$ is

$$\begin{aligned}\ln q(\beta_t) &= \mathbb{E}_{q(\Theta\setminus\beta_t)}[\ln p(\boldsymbol{\mathcal{X}}_{\boldsymbol{\Omega}}, \Theta)] \\ &= \mathbb{E}\left[\ln \beta_t + \alpha \ln(1-\beta_t) + \sum_{i=1}^{I} q(z_i > t)\ln(1-\beta_t) + q(z_i = t)\ln(\beta_t)\right] + \text{const} \\ &= \mathbb{E}\left[\ln \beta_t + \alpha \ln(1-\beta_t) + \sum_{i=1}^{I}\sum_{m=t+1}^{T} \phi_{im}\ln(1-\beta_t) + \phi_{it}\ln(\beta_t)\right] + \text{const}\end{aligned}$$

The above expression can be simplified as $q(\beta_t) = \text{Beta}(\gamma_{t1}, \gamma_{t2})$, where

$$\begin{aligned}\gamma_{t1} &= 1 + \sum_{i=1}^{I} \phi_{it} \\ \gamma_{t2} &= \alpha + \sum_{i=1}^{I}\sum_{m=t+1}^{T} \phi_{im}\end{aligned}$$

*3. The variational posterior distribution of* $\{\boldsymbol{\mu}_t\}_{t=1}^{T}$

The optimal logarithmic variational posterior distribution of $\boldsymbol{\mu}_t$ is

$$\begin{aligned}\ln q(\boldsymbol{\mu}_t) &= \mathbb{E}_{q(\Theta\setminus\boldsymbol{\mu}_t)}[\ln p(\boldsymbol{\mathcal{X}}_{\boldsymbol{\Omega}}, \Theta)] \\ &= \mathbb{E}\left[-\frac{1}{2}\sum_{i=1}^{I} \phi_{it}(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t)^T(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t) - \frac{\lambda}{2}\boldsymbol{\mu}_t^T\boldsymbol{\mu}_t\right] + \text{const}\end{aligned}$$

The above expression can be simplified as $q(\boldsymbol{\mu}_t) = \mathcal{N}\left(\tilde{\boldsymbol{\mu}}_t, \left[\tilde{\boldsymbol{\Lambda}}_t\right]^{-1}\right)$, where

$$\begin{aligned}\tilde{\boldsymbol{\Lambda}}_t &= \left(\sum_{i=1}^{I}\phi_{it} + \lambda\right) I \\ \boldsymbol{\mu}_t &= \left[\tilde{\boldsymbol{\Lambda}}_t\right]^{-1}\left(\sum_{i=1}^{I}\phi_{it}\boldsymbol{u}_{i:}\right)\end{aligned}$$

*4. The variational posterior distribution of* $(\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v)$

The optimal varitional posterior distribution of $(\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v)$ is

$$\begin{aligned}
\ln q\left(\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v\right) &= \mathbb{E}_{q(\Theta \backslash \boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v)}\left[\ln p\left(\boldsymbol{\mathcal{X}}_{\boldsymbol{\Omega}}, \Theta\right)\right] \\
&= \mathbb{E}\left[\frac{J}{2}\ln\left|\boldsymbol{\Lambda}_v\right| - \frac{1}{2}\sum_{j=1}^{J}\left(\boldsymbol{v}_{j:} - \boldsymbol{\mu}_v\right)^T \boldsymbol{\Lambda}_v\left(\boldsymbol{v}_{j:} - \boldsymbol{\mu}_v\right) + \frac{1}{2}\ln\left|\boldsymbol{\Lambda}_v\right| - \frac{\beta_0}{2}\left(\boldsymbol{\mu}_v - \boldsymbol{\mu}_0\right)^T \boldsymbol{\Lambda}_v\left(\boldsymbol{\mu}_v - \boldsymbol{\mu}_0\right)\right. \\
&\quad \left. + \frac{v_0 - R - 1}{2}\ln\left|\boldsymbol{\Lambda}_v\right| - \frac{1}{2}\mathrm{trace}\left(\boldsymbol{W}_0^{-1}\boldsymbol{\Lambda}_v\right)\right] + \mathrm{const} \\
&= \mathbb{E}\left[\frac{1}{2}\ln\left|\boldsymbol{\Lambda}_v\right| - \frac{J+\beta_0}{2}\left(\boldsymbol{\mu}_v - \tilde{\boldsymbol{\mu}}_v\right)^T \boldsymbol{\Lambda}_v\left(\boldsymbol{\mu}_v - \tilde{\boldsymbol{\mu}}_v\right) + \frac{J + v_0 - R - 1}{2}\ln\left|\boldsymbol{\Lambda}_v\right|\right. \\
&\quad \left. + \frac{1}{2}\mathrm{trace}\left(\left[\boldsymbol{W}_0^{-1} + \sum_{j=1}^{J}\left(\boldsymbol{v}_{j:} - \bar{\boldsymbol{v}}\right)\boldsymbol{v}_{j:}^T + \frac{J\beta_0}{J+\beta_0}\left(\bar{\boldsymbol{v}} - \boldsymbol{\mu}_0\right)\left(\bar{\boldsymbol{v}} - \boldsymbol{\mu}_0\right)^T\right]\boldsymbol{\Lambda}_v\right)\right] + \mathrm{const}
\end{aligned}$$

The variational posterior distribution of $(\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v)$ is $q\left(\boldsymbol{\mu}_v, \boldsymbol{\Lambda}_v\right) = \mathcal{N}\left(\tilde{\boldsymbol{\mu}}_v, \left[\tilde{\boldsymbol{\Lambda}}_v\right]^{-1}\right) \times \mathrm{Wishart}\left(\tilde{\boldsymbol{W}}_v, \tilde{v}_v\right)$.

$$\begin{aligned}
\bar{\boldsymbol{v}} &= \frac{\sum_{j=1}^{J}\boldsymbol{v}_j}{J}, \quad \tilde{\boldsymbol{\mu}}_v = \frac{\sum_{j=1}^{J}\mathbb{E}\left[\boldsymbol{v}_j\right] + \beta_0\boldsymbol{\mu}_0}{J+\beta_0}, \quad \tilde{\boldsymbol{\Lambda}}_v = \left(J+\beta_0\right)\mathbb{E}\left[\boldsymbol{\Lambda}_v\right] \\
\tilde{\boldsymbol{W}}_v^{-1} &= \mathbb{E}\left[\boldsymbol{W}_0^{-1} + \sum_{j=1}^{J}\left(\boldsymbol{v}_j - \bar{\boldsymbol{v}}\right)\boldsymbol{v}_j^T + \frac{J\beta_0}{J+\beta_0}\left(\bar{\boldsymbol{v}} - \boldsymbol{\mu}_0\right)\left(\bar{\boldsymbol{v}} - \boldsymbol{\mu}_0\right)^T\right], \quad \tilde{v}_v = J + v_0
\end{aligned}$$

Similarly, we can derive the variational posterior distribution of $(\boldsymbol{\mu}_w, \boldsymbol{\Lambda}_w)$.

*5. The variational posterior distribution of* $\boldsymbol{U}$

Take inference $i$-th latent vector $\boldsymbol{u}_{i:}$ of factor matrix $\boldsymbol{U}$ as a example. The optimal form of posterior distribution is

$$\begin{aligned}
\ln q\left(\boldsymbol{u}_{i:}\right) &= \mathbb{E}_{q(\Theta \backslash \boldsymbol{u}_{i:})}\left[\ln p\left(\boldsymbol{\mathcal{X}}_{\boldsymbol{\Omega}}, \Theta\right)\right] + \mathrm{const} \\
&= \mathbb{E}\left[-\frac{\tau}{2}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(o_{ijk}\left(x_{ijk} - \langle \boldsymbol{u}_{i:}, \boldsymbol{v}_{j:}, \boldsymbol{w}_{k:}\rangle\right)^2\right) - \frac{1}{2}\sum_{t=1}^{T}\phi_{it}\left(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t\right)^T \boldsymbol{\Lambda}_t\left(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t\right)\right] + \mathrm{const} \\
&= \mathbb{E}\left[-\frac{\tau}{2}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(o_{ijk}\left(x_{ijk} - \boldsymbol{u}_{i:}^T\left(\boldsymbol{v}_{j:} \circledast \boldsymbol{w}_{k:}\right)\right)^2\right) - \frac{1}{2}\sum_{t=1}^{T}\phi_{it}\left(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t\right)^T \boldsymbol{\Lambda}_t\left(\boldsymbol{u}_{i:} - \boldsymbol{\mu}_t\right)\right] + \mathrm{const}
\end{aligned}$$

The above expression can be simplified as $q\left(\boldsymbol{u}_i\right) = \mathcal{N}\left(\tilde{\boldsymbol{u}}_i, \left[\tilde{\boldsymbol{\Lambda}}_{u_i}\right]^{-1}\right)$, where

$$\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{u_i} &= \mathbb{E}\left[\tau\sum_{j=1}^{J}\sum_{k=1}^{K}\left(o_{ijk}\left(\boldsymbol{v}_{j:} \circledast \boldsymbol{w}_{k:}\right)\left(\boldsymbol{v}_{j:} \circledast \boldsymbol{w}_{k:}\right)^T\right) + \sum_{t=1}^{T}\phi_{it}\boldsymbol{\Lambda}_t\right] \\
\tilde{\boldsymbol{u}}_i &= \left[\tilde{\boldsymbol{\Lambda}}_{u_i}\right]^{-1}\mathbb{E}\left[\tau\sum_{j=1}^{J}\sum_{k=1}^{K}\left(x_{ijk}\left(\boldsymbol{v}_{j:} \circledast \boldsymbol{w}_{k:}\right)\right) + \sum_{t=1}^{T}\boldsymbol{\Lambda}_t\boldsymbol{\mu}_t\right]
\end{aligned}$$

*6. The variational posterior distribution of $\boldsymbol{V}$*

Take inference $j$-th latent vector $\boldsymbol{v}_{j:}$ of factor matrix $\boldsymbol{V}$ as a example. The optimal form of variational posterior distribution is

$$\begin{aligned}
\ln q\left(\boldsymbol{v}_{j:}\right) &= \mathbb{E}_{q(\Theta \backslash \boldsymbol{v}_{j:})}\left[\ln p\left(\boldsymbol{\mathcal{X}}_{\boldsymbol{\Omega}}, \Theta\right)\right] + \text{const} \\
&= \mathbb{E}\left[-\frac{\tau}{2}\sum_{i=1}^{I}\sum_{k=1}^{K}\left(o_{ijk}\left(x_{ijk} - \langle \boldsymbol{u}_{i:}, \boldsymbol{v}_{j:}, \boldsymbol{w}_{k:}\rangle\right)^2\right) - \frac{1}{2}\left(\boldsymbol{v}_{j:} - \boldsymbol{\mu}_v\right)^T \boldsymbol{\Lambda}_v \left(\boldsymbol{v}_{j:} - \boldsymbol{\mu}_v\right)\right] + \text{const} \\
&= \mathbb{E}\left[-\frac{\tau}{2}\sum_{i=1}^{I}\sum_{k=1}^{K}\left(o_{ijk}\left(x_{ijk} - \boldsymbol{u}_{i:}^T\left(\boldsymbol{v}_{j:} \circledast \boldsymbol{w}_{k:}\right)\right)^2\right) - \frac{1}{2}\left(\boldsymbol{v}_{j:} - \boldsymbol{\mu}_v\right)^T \boldsymbol{\Lambda}_v \left(\boldsymbol{v}_{j:} - \boldsymbol{\mu}_v\right)\right] + \text{const}
\end{aligned}$$

The above expression can be simplified as $q\left(\boldsymbol{v}_{j:}\right) = \mathcal{N}\left(\tilde{\boldsymbol{v}}_j, \left[\tilde{\boldsymbol{\Lambda}}_{v_j}\right]^{-1}\right)$, where

$$\begin{aligned}
\tilde{\boldsymbol{\Lambda}}_{v_j} &= \mathbb{E}\left[\tau\sum_{i=1}^{I}\sum_{k=1}^{K}\left(o_{ijk}\left(\boldsymbol{u}_{i:} \circledast \boldsymbol{w}_{k:}\right)\left(\boldsymbol{u}_{i:} \circledast \boldsymbol{w}_{k:}\right)^T\right) + \boldsymbol{\Lambda}_v\right] \\
\tilde{\boldsymbol{v}}_j &= \left[\tilde{\boldsymbol{\Lambda}}_{v_j}\right]^{-1}\mathbb{E}\left[\tau\sum_{i=1}^{I}\sum_{k=1}^{K}\left(x_{ijk}\left(\boldsymbol{u}_{i:} \circledast \boldsymbol{w}_{k:}\right)\right) + \boldsymbol{\Lambda}_v\boldsymbol{\mu}_v\right]
\end{aligned}$$

Similarly, we can derive the variational posterior distribution of $\boldsymbol{W}$.

*7. The variational posterior distribution of $\tau$*

The optimal logarithmic variational posterior distribution of $\tau$ is

$$\begin{aligned}
\ln q(\tau) &= \mathbb{E}_{q(\Theta \backslash \tau)}\left[\ln p\left(\boldsymbol{\mathcal{X}}_{\boldsymbol{\Omega}}, \Theta\right)\right] \\
&= \mathbb{E}\left[\sum_{(i,j,k)\in\Omega}\left(\frac{1}{2}\ln\tau - \frac{\tau}{2}\left(x_{ijk} - \langle \boldsymbol{u}_{i:}, \boldsymbol{v}_{j:}, \boldsymbol{w}_{k:}\rangle\right)^2\right) + (a_0 - 1)\ln\tau - b_0\tau\right] + \text{const} \\
&= \left(\frac{1}{2}\sum_{(i,j,k)\in\Omega} o_{ijk} + a_0 - 1\right)\ln\tau - \left(\mathbb{E}\left[\sum_{(i,j,k)\in\Omega}\frac{1}{2}\left(x_{ijk} - \langle \boldsymbol{u}_{i:}, \boldsymbol{v}_{j:}, \boldsymbol{w}_{k:}\rangle\right)^2\right] + b_0\right)\tau + \text{const}
\end{aligned}$$

The above expression can be simplified as $q(\tau) = \text{Gamma}\left(\tilde{a}_0, \tilde{b}_0\right)$, where

$$\begin{aligned}
\tilde{a}_0 &= \frac{1}{2}\sum_{(i,j,k)\in\Omega} o_{ijk} + a_0 \\
\tilde{b}_0 &= \mathbb{E}\left[\sum_{(i,j,k)\in\Omega}\left(x_{ijk} - \langle \boldsymbol{u}_{i:}, \boldsymbol{v}_{j:}, \boldsymbol{w}_{k:}\rangle\right)^2\right] + b_0
\end{aligned}$$