Supplementary Materials for "Probabilistic Rank-One Tensor Analysis with Concurrent Regularizations"

1 Derivations of the Log-Likelihood Function

The detailed derivations of the log-likelihood function, i.e., expression (6) in our paper, are as follows: The p.d.f. of the matrix-variate distribution $p(\mathbf{X}) = \mathcal{N}_{I_1,I_2}(\mathbf{X}|\mathbf{\Xi}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ is given by

$$p(\mathbf{X}) = (2\pi)^{-\frac{1}{2}I_1 I_2}|\mathbf{\Sigma}_1|^{-\frac{1}{2}I_2}|\mathbf{\Sigma}_2|^{-\frac{1}{2}I_1}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}_1^{-1}(\mathbf{X}-\mathbf{\Xi})\mathbf{\Sigma}_2^{-1}(\mathbf{X}-\mathbf{\Xi})^\top\right)\right\}. \quad (1)$$

With the above results, the conditional distribution $p(\mathbf{X}_{m(n)}|\mathbf{z}_m) = \mathcal{N}_{I_n,I^{(n^-)}}(\mathbf{X}_{m(n)}|\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top}, \sigma\mathbf{I}_{I_n}, \sigma\mathbf{I}_{I^{(n^-)}})$, i.e., expression (5) in our paper, can be written as follows:

$$\begin{aligned}
p(\mathbf{X}_{m(n)}|\mathbf{z}_m) &= (2\pi)^{-\frac{1}{2}I_n I^{(n^-)}}|\sigma\mathbf{I}_{I_n}|^{-\frac{1}{2}I^{(n^-)}}|\sigma\mathbf{I}_{I^{(n^-)}}|^{-\frac{1}{2}I_n} \\
&\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\sigma^{-1}\mathbf{I}_{I_n}(\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top})\sigma^{-1}\mathbf{I}_{I^{(n^-)}}(\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top})^\top\right)\right\} \\
&= (2\pi)^{-\frac{1}{2}I_n I^{(n^-)}}\sigma^{-\frac{1}{2}I^{(n^-)}I_n}\sigma^{-\frac{1}{2}I_n I^{(n^-)}}\exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top}\|_F^2\right\} \\
&= (2\pi)^{-\frac{1}{2}I}(\sigma^2)^{-\frac{1}{2}I}\exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top}\|_F^2\right\}, \quad (2)
\end{aligned}$$

where $\mathbf{U}^{(n^-)} \in \mathbb{R}^{I^{(n^-)}\times P}$ is the mode-n *complement* factor matrix with $\mathbf{U}^{(n^-)} = \mathbf{U}^{(N)} \odot \ldots \odot \mathbf{U}^{(n+1)} \odot \mathbf{U}^{(n-1)} \odot \ldots \odot \mathbf{U}^{(1)}$, $I^{(n^-)} = \prod_{k\neq n} I_k$, and $I = I_n I^{(n^-)} = \prod_n I_n$. The above p.d.f. is obtained by substituting $\mathbf{X} = \mathbf{X}_{m(n)}$, $\mathbf{\Xi} = \mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top}$, $\mathbf{\Sigma}_1 = \sigma\mathbf{I}_{I_n}$, and $\mathbf{\Sigma}_2 = \sigma\mathbf{I}_{I^{(n^-)}}$ into (1). Here, we have used the subscript to explicitly indicate the sizes of the *identity* matrices $\mathbf{I}_{I_n} \in \mathbb{R}^{I_n\times I_n}$ and $\mathbf{I}_{I^{(n^-)}} \in \mathbb{R}^{I^{(n^-)}\times I^{(n^-)}}$ for clarity.

From (2), the logarithm of $p(\mathbf{X}_{m(n)}|\mathbf{z}_m)$ is given by

$$\ln p(\mathbf{X}_{m(n)}|\mathbf{z}_m) = -\frac{1}{2}\{I\ln 2\pi + I\ln\sigma^2 + \frac{1}{\sigma^2}\|\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top}\|_F^2\}. \quad (3)$$

In addition, $p(\mathbf{z}_m) = \mathcal{N}(\mathbf{z}_m|\mathbf{0},\mathbf{I}) = (2\pi)^{-\frac{1}{2}P}\exp\{-\frac{1}{2}\mathbf{z}_m^\top\mathbf{z}_m\}$, and its logarithm is given by $\ln p(\mathbf{z}_m) = -\frac{1}{2}\{P\ln 2\pi + \mathbf{z}_m^\top\mathbf{z}_m\}$. Therefore, the expectation of the log-likelihood function, i.e., expression (6) in our paper, can be derived as follows:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \sum_{m=1}^{M}\langle\ln p(\mathbf{X}_{m(n)},\mathbf{z}_m)\rangle = \sum_{m=1}^{M}\langle\ln p(\mathbf{X}_{m(n)}|\mathbf{z}_m) + \ln p(\mathbf{z}_m)\rangle \\
&= -\frac{1}{2}\sum_{m=1}^{M}\left\{\{I\ln 2\pi + I\ln\sigma^2 + \frac{1}{\sigma^2}\langle\|\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z})\mathbf{U}^{(n^-)^\top}\|_F^2\rangle\} + \{P\ln 2\pi + \langle\mathbf{z}_m^\top\mathbf{z}_m\rangle\}\right\} \\
&= -\sum_{m=1}^{M}\left[\frac{I}{2}\ln\sigma^2 + \frac{1}{2}\langle\mathbf{z}_m^\top\mathbf{z}_m\rangle + \frac{1}{2\sigma^2}\langle||\mathbf{X}_{m(n)}-\mathbf{U}^{(n)}\mathrm{diag}(\mathbf{z}_m)\mathbf{U}^{(n^-)^\top}||_F^2\rangle\right] + \mathrm{const}. \quad (4)
\end{aligned}$$

2 Derivations of the Update of $\mathbf{U}^{(n)}$

Given the parameter set $\boldsymbol{\theta} = \{\mathbf{U}^{(n)}, \mathbf{U}^{(n^-)}, \sigma^2\}$, the log-likelihood function (4) can be rewritten by grouping the terms related to $\mathbf{U}^{(n)}$ together and *omitting* other terms as a constant, which leads to

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= -\frac{1}{2\sigma^2}\sum_{m=1}^{M}\langle||\mathbf{X}_{m(n)} - \mathbf{U}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle \\
&= -\frac{1}{2\sigma^2}\sum_{m=1}^{M}\text{tr}\left(\langle(\mathbf{X}_{m(n)} - \mathbf{U}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top)(\mathbf{X}_{m(n)} - \mathbf{U}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top)^\top\rangle\right) \\
&= -\frac{1}{2\sigma^2}\sum_{m=1}^{M}\text{tr}\left(\mathbf{X}_{m(n)}{\mathbf{X}_{m(n)}}^\top + \langle\mathbf{U}^{(n)}\text{diag}(\mathbf{z}){\mathbf{U}^{(n^-)}}^\top\mathbf{U}^{(n^-)}\text{diag}(\mathbf{z}){\mathbf{U}^{(n)}}^\top\rangle - 2\mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle){\mathbf{U}^{(n)}}^\top\right) \\
&= -\frac{1}{2\sigma^2}\sum_{m=1}^{M}\text{tr}\left(\mathbf{X}_{m(n)}{\mathbf{X}_{m(n)}}^\top + \mathbf{U}^{(n)}(\langle\mathbf{z}_m\mathbf{z}_m^\top\rangle \circledast {\mathbf{U}^{(n^-)}}^\top\mathbf{U}^{(n^-)}){\mathbf{U}^{(n)}}^\top - 2\mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle){\mathbf{U}^{(n)}}^\top\right). \quad (5)
\end{aligned}$$

Then, we can take the partial derivative with respect to $\mathbf{U}^{(n)}$ and solve

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\mathbf{U}^{(n)}} = -\frac{1}{2\sigma^2}\sum_{m=1}^{M}\left\{2\mathbf{U}^{(n)}(\langle\mathbf{z}_m\mathbf{z}_m^\top\rangle \circledast {\mathbf{U}^{(n^-)}}^\top\mathbf{U}^{(n^-)}) - 2\mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle)\right\} = 0. \quad (6)$$

It is clear that the solution of $\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\mathbf{U}^{(n)}}$ is given by

$$\begin{aligned}
\tilde{\mathbf{U}}^{(n)}\left[\sum_{m=1}^{M}\langle\mathbf{z}_m\mathbf{z}_m^\top\rangle \circledast {\mathbf{U}^{(n^-)}}^\top\mathbf{U}^{(n^-)}\right] &= \sum_{m=1}^{M}\mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle) \\
\tilde{\mathbf{U}}^{(n)} &= \left[\sum_{m=1}^{M}\mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle)\right]\left[\sum_{m=1}^{M}\langle\mathbf{z}_m\mathbf{z}_m^\top\rangle \circledast {\mathbf{U}^{(n^-)}}^\top\mathbf{U}^{(n^-)}\right]^{-1}, \quad (7)
\end{aligned}$$

which leads to expression (11) in our paper.

3 Derivations of the Update of σ^2

Similarly, we can rewrite the log-likelihood function (4) by only considering the terms related to σ^2, which leads to

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{MI}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{m=1}^{M}\langle||\mathbf{X}_{m(n)} - \mathbf{U}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle. \quad (8)$$

We then take the partial derivative with respect to σ^2 and solve

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\sigma^2} = -\frac{MI}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{m=1}^{M}\langle||\mathbf{X}_{m(n)} - \mathbf{U}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle = 0. \quad (9)$$

The solution of $\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\sigma^2} = 0$ is given by

$$\begin{aligned}
\frac{MI}{2}\frac{1}{\tilde{\sigma}^2} &= \frac{1}{2(\tilde{\sigma}^2)^2}\sum_{m=1}^{M}\langle||\mathbf{X}_{m(n)} - \tilde{\mathbf{U}}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle \\
MI\tilde{\sigma}^2 &= \sum_{m=1}^{M}\langle||\mathbf{X}_{m(n)} - \tilde{\mathbf{U}}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle \\
\tilde{\sigma}^2 &= \frac{1}{MI}\sum_{m=1}^{M}\langle||\mathbf{X}_{m(n)} - \tilde{\mathbf{U}}^{(n)}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle, \quad (10)
\end{aligned}$$

where the solution of $\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\mathbf{U}^{(n)}} = 0$, i.e. $\tilde{\mathbf{U}}^{(n)}$ (updated $\mathbf{U}^{(n)}$), has been used to compute $\langle||\mathbf{X}_{m(n)} - \tilde{\mathbf{U}}\text{diag}(\mathbf{z}_m){\mathbf{U}^{(n^-)}}^\top||_F^2\rangle$.

The above solution of σ^2 can be further simplified by substituting (7) into (10). From (5), we have

$$\begin{aligned}
\tilde{\sigma}^2 &= \frac{1}{MI}\sum_{m=1}^{M}\text{tr}\left(\mathbf{X}_{m(n)}{\mathbf{X}_{m(n)}}^\top + \tilde{\mathbf{U}}^{(n)}(\langle\mathbf{z}_m\mathbf{z}_m^\top\rangle \circledast {\mathbf{U}^{(n^-)}}^\top\mathbf{U}^{(n^-)})\tilde{\mathbf{U}}^{(n)\top} - 2\mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle)\tilde{\mathbf{U}}^{(n)\top}\right) \\
&= \frac{1}{MI}\sum_{m=1}^{M}\text{tr}\left(\mathbf{X}_{m(n)}{\mathbf{X}_{m(n)}}^\top - \mathbf{X}_{m(n)}\mathbf{U}^{(n^-)}\text{diag}(\langle\mathbf{z}_m\rangle)\tilde{\mathbf{U}}^{(n)\top}\right), \quad (11)
\end{aligned}$$

which leads to expression (12) in our paper. It is worth noting that the second equality of (11) holds because of the following fact:

$$\sum_{m=1}^{M} \tilde{\mathbf{U}}^{(n)}(\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)}) \tilde{\mathbf{U}}^{(n)\top} = \tilde{\mathbf{U}}^{(n)} \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} \right] \tilde{\mathbf{U}}^{(n)\top}$$
$$= \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle) \right] \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} \right]^{-1} \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} \right] \tilde{\mathbf{U}}^{(n)\top}$$
$$= \sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle) \tilde{\mathbf{U}}^{(n)\top}. \tag{12}$$

4 Derivations of the Update of $\mathbf{U}^{(n)}$ with L_2 Regularization

Recall that the L_2-regularized log-likelihood function is

$$\mathcal{L}^{L_2}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) - \gamma \sum_{n=1}^{N} \text{tr}(\mathbf{U}^{(n)} {\mathbf{U}^{(n)}}^\top), \tag{13}$$

where $\mathcal{L}(\boldsymbol{\theta})$ is given by (4). By only considering the terms related to $\mathbf{U}^{(n)}$, $\mathcal{L}^{L_2}(\boldsymbol{\theta})$ becomes

$$\mathcal{L}^{L_2}(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_{m=1}^{M} \langle || \mathbf{X}_{m(n)} - \mathbf{U}^{(n)} \text{diag}(\mathbf{z}_m) {\mathbf{U}^{(n^-)}}^\top ||_F^2 \rangle - \gamma \text{tr}(\mathbf{U}^{(n)} {\mathbf{U}^{(n)}}^\top). \tag{14}$$

With the similar derivations in (6), we take the partial derivative of $\mathcal{L}^{L_2}(\boldsymbol{\theta})$ with respect to $\mathbf{U}^{(n)}$ and solve

$$\frac{\partial \mathcal{L}^{L_2}(\boldsymbol{\theta})}{\partial \mathbf{U}^{(n)}} = -\frac{1}{2\sigma^2} \sum_{m=1}^{M} \left\{ 2\mathbf{U}^{(n)}(\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)}) - 2\mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle) \right\} - 2\gamma \mathbf{U}^{(n)} = 0. \tag{15}$$

The solution of $\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mathbf{U}^{(n)}} = 0$ can be obtained by

$$\frac{1}{\sigma^2} \tilde{\mathbf{U}}^{(n)} \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} \right] + 2\gamma \tilde{\mathbf{U}}^{(n)} = \frac{1}{\sigma^2} \sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)$$
$$\tilde{\mathbf{U}}^{(n)} \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} \right] + 2\sigma^2 \gamma \tilde{\mathbf{U}}^{(n)} = \sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)$$
$$\tilde{\mathbf{U}}^{(n)} \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} + 2\sigma^2 \gamma \mathbf{I} \right] = \sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)$$
$$\tilde{\mathbf{U}}^{(n)} = \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle) \right] \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} + 2\sigma^2 \gamma \mathbf{I} \right]^{-1}. \tag{16}$$

Since σ^2 is considered as a constant during the update of $\mathbf{U}^{(n)}$, without loss of generality, the scale $2\sigma^2$ can be *absorbed into* the regularization parameter γ, and finally we have

$$\tilde{\mathbf{U}}^{(n)} = \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle) \right] \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} + \gamma \mathbf{I} \right]^{-1}, \tag{17}$$

which leads to expression (14) in our paper.

5 Derivations of the Update of $\mathbf{U}^{(n)}$ with Moment-Based Concurrent Regularization

Recall that moment-based CR aims to improve the conditioning of $\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle$ as follows:

$$\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle^{\text{MCR}} = \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle + \frac{\gamma}{M} \mathbf{I}. \tag{18}$$

Replacing the original second-order moment $\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle$ by $\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle^{\text{MCR}}$ in (7), we have

$$\begin{aligned}
\tilde{\mathbf{U}}^{(n)} &= \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)\right] \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle^{\text{MCR}} \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)}\right]^{-1} \\
&= \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)\right] \left[\sum_{m=1}^{M} (\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle + \frac{\gamma}{M}\mathbf{I}) \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)}\right]^{-1} \\
&= \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)\right] \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} + \gamma \mathbf{I} \circledast ({\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)})\right]^{-1} \\
&= \left[\sum_{m=1}^{M} \mathbf{X}_{m(n)} \mathbf{U}^{(n^-)} \text{diag}(\langle \mathbf{z}_m \rangle)\right] \left[\sum_{m=1}^{M} \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle \circledast {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} + \gamma \mathbf{\Lambda}^{(n^-)}\right]^{-1},
\end{aligned} \tag{19}$$

where we have defined $\mathbf{\Lambda}^{(n^-)} = \mathbf{I} \circledast ({\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)})$. This eventually leads to expression (17) in our paper.

6 Derivations of the Joint Distribution for PROTA with Bayesian CR

PROTA with Bayesian CR has the following joint distribution:

$$p(\mathcal{D}, \mathbf{\Theta}) = \prod_{m=1}^{M} p(\mathcal{X}_m | \mathbf{z}_m, \{\mathbf{U}^{(n)}\}_{n=1}^{N}, \tau) \prod_{m=1}^{M} p(\mathbf{z}_m) \prod_{n=1}^{N} p(\mathbf{U}^{(n)}) p(\tau). \tag{20}$$

The logarithm of $p(\mathcal{D}, \mathbf{\Theta})$ is given by:

$$\begin{aligned}
\ln p(\mathcal{D}, \mathbf{\Theta}) = &-\frac{1}{2} \sum_{m=1}^{M} \left[\tau ||\mathbf{X}_{m(n)} - \mathbf{U}^{(n)} \text{diag}(\mathbf{z}_m) {\mathbf{U}^{(n^-)}}^\top ||_F^2 - I \ln \tau + \mathbf{z}_m^\top \mathbf{z}_m\right] \\
&- \frac{1}{2}\gamma \langle \tau \rangle \text{tr}(\sum_{n=1}^{N} \langle \mathbf{\Lambda}^{(\backslash n)} \rangle \mathbf{U}^{(n)\top} \mathbf{U}^{(n)}) + (a_0 - 1)\ln \tau - b_0 \tau + \text{const}.
\end{aligned} \tag{21}$$

With the above formulation, we can perform variational inference by substituting (21) into the optimized form of the variational distributions as follows:

$$\ln q_j(\mathbf{\Theta}_j) \propto \langle \ln p(\mathcal{D}, \mathbf{\Theta}) \rangle_{\mathbf{\Theta} \backslash \mathbf{\Theta}_j}. \tag{22}$$

7 Expectations for the Variational Updates

The expectations involved in updating the variational distributions $q(\mathbf{z}_m)$, $q(\mathbf{U}^{(n)})$, and $q(\tau)$, i.e., equations (24), (25), and (26) in our paper, respectively, can be computed as follows:

$$\langle \tau \rangle = \frac{a_\tau}{b_\tau}, \tag{23}$$

$$\langle \mathbf{U}^{(n)} \rangle = \sum_{m=1}^{M} \mathbf{X}_{m(n)} \langle \mathbf{U}^{(n^-)} \rangle \text{diag}(\langle \mathbf{z}_m \rangle) \mathbf{\Sigma}^{(n)}, \tag{24}$$

$$\langle {\mathbf{U}^{(n)}}^\top \mathbf{U}^{(n)} \rangle = I_n \mathbf{\Sigma}^{(n)} + \langle \mathbf{U}^{(n)} \rangle^\top \langle \mathbf{U}^{(n)} \rangle, \tag{25}$$

$$\langle \mathbf{U}^{(n^-)} \rangle = \langle \mathbf{U}^{(N)} \rangle \odot \ldots \odot \langle \mathbf{U}^{(n+1)} \rangle \odot \langle \mathbf{U}^{(n-1)} \rangle \odot \ldots \odot \langle \mathbf{U}^{(1)} \rangle, \tag{26}$$

$$\langle {\mathbf{U}^{(n^-)}}^\top \mathbf{U}^{(n^-)} \rangle = \circledast_{k \neq n} \langle {\mathbf{U}^{(k)}}^\top \mathbf{U}^{(k)} \rangle \tag{27}$$

$$\langle \mathbf{W} \rangle = \langle \mathbf{U}^{(N)} \rangle \odot \ldots \odot \langle \mathbf{U}^{(1)} \rangle, \tag{28}$$

$$\langle \mathbf{W}^\top \mathbf{W} \rangle = \circledast_{n=1}^{N} \langle {\mathbf{U}^{(n)}}^\top \mathbf{U}^{(n)} \rangle, \tag{29}$$

$$\langle \mathbf{z}_m \rangle = \langle \tau \rangle \mathbf{\Sigma}_{\mathbf{z}} \langle \mathbf{W} \rangle^\top \text{vec}(\mathcal{X}_m), \tag{30}$$

$$\langle \mathbf{z}_m \mathbf{z}_m^\top \rangle = \mathbf{\Sigma}_{\mathbf{z}} + \langle \mathbf{z}_m \rangle \langle \mathbf{z}_m \rangle^\top, \tag{31}$$

where we define $\circledast_{n=1}^{N} \langle {\mathbf{U}^{(n)}}^\top \mathbf{U}^{(n)} \rangle = \langle {\mathbf{U}^{(N)}}^\top \mathbf{U}^{(N)} \rangle \circledast \ldots \circledast \langle {\mathbf{U}^{(1)}}^\top \mathbf{U}^{(1)} \rangle$, and $\circledast_{k \neq n} \langle {\mathbf{U}^{(k)}}^\top \mathbf{U}^{(k)} \rangle = \langle {\mathbf{U}^{(N)}}^\top \mathbf{U}^{(N)} \rangle \circledast \ldots \circledast \langle {\mathbf{U}^{(n+1)}}^\top \mathbf{U}^{(n+1)} \rangle \circledast \langle {\mathbf{U}^{(n-1)}}^\top \mathbf{U}^{(n-1)} \rangle \circledast \ldots \circledast \langle {\mathbf{U}^{(1)}}^\top \mathbf{U}^{(1)} \rangle$. Finally, the expectation of the model fitting error can be computed by

$$\langle ||\text{vec}(\mathcal{X}_m) - \mathbf{W}\mathbf{z}_m||^2 \rangle = \text{vec}(\mathcal{X}_m)^\top \text{vec}(\mathcal{X}_m) - 2\text{vec}(\mathcal{X}_m) \langle \mathbf{W} \rangle \langle \mathbf{z}_m \rangle + \text{tr}(\langle \mathbf{W}^\top \mathbf{W} \rangle \langle \mathbf{z}_m \mathbf{z}_m^\top \rangle). \tag{32}$$