Supplementary Materials of "Block Tensor Ring Decomposition: Theory and Application"

Sheng Liu, Xi-Le Zhao*, and Hao Zhang

I. THE PROOF OF THEOREM 1

Before proving Theorem 1, we need to introduce and establish some lemmas.

Lemma 1 (permutation lemma [1]). *Consider two matrices $\bar{\mathbf{A}}, \mathbf{A} \in \mathbb{R}^{n\times d}$, that have no zero columns. If for every vector $\mathbf{x}$ such that $\omega\left(\mathbf{x}^\top \bar{\mathbf{A}}\right) \leqslant d - r_{\bar{\mathbf{A}}} + 1$, we have $\omega\left(\mathbf{x}^T \mathbf{A}\right) \leqslant \omega\left(\mathbf{x}^\top \bar{\mathbf{A}}\right)$, then there exists a unique permutation matrix $\mathbf{\Pi}$ and a unique nonsingular diagonal matrix $\mathbf{\Lambda}$ such that $\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}\mathbf{\Lambda}$.*

Lemma 2. *For $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3\times n_4}$, its BTRD is $\mathcal{X} = \sum_{k=1}^{d} \boldsymbol{e}^k \circ TR\left(\mathcal{F}_1^k, \mathcal{F}_2^k, \mathcal{F}_3^k\right)$ with $R_1 = R_2 = R$ and $R_3 = 1$. Suppose $\boldsymbol{F}_1, \boldsymbol{F}_2$, and $\boldsymbol{F}_3$ are full column rank matrices satisfy*

$$k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} + k'_{\boldsymbol{E}} + \geq 2d + 3, \tag{1}$$

where $k'_{\boldsymbol{F}_1}$ is k'-rank (Kruskal rank) of $\boldsymbol{F}_1$, $\boldsymbol{F}_1 = \left[\left(\mathcal{F}_1^1\right)_{(3)}^\top, \cdots, \left(\mathcal{F}_1^d\right)_{(3)}^\top\right]$, $\boldsymbol{F}_2 = \left[\left(\mathcal{F}_2^1\right)_{(1)}^\top, \cdots, \left(\mathcal{F}_2^d\right)_{(1)}^\top\right]$, $\boldsymbol{F}_3 = \left[\left(\mathcal{F}_3^1\right)_{(1)}^\top, \cdots, \left(\mathcal{F}_3^d\right)_{(1)}^\top\right]$ and $\boldsymbol{E} = [\boldsymbol{e}^1, \cdots, \boldsymbol{e}^d]$. Then, we have an alternative decomposition of $\mathcal{X}$, i.e., $\mathcal{X} = \sum_{k=1}^{d} \bar{\boldsymbol{e}}^k \circ TR\left(\bar{\mathcal{F}}_1^k, \bar{\mathcal{F}}_2^k, \bar{\mathcal{F}}_3^k\right)$, where there holds $\bar{\boldsymbol{E}} = \boldsymbol{E}\Pi_{\boldsymbol{E}}\Lambda_{\boldsymbol{E}}$, $\Pi_{\boldsymbol{E}}$ is a permutation matrix and $\Lambda_{\boldsymbol{E}}$ a nonsingular diagonal matrix.

Proof. Based on Lemma 1, for any $\boldsymbol{y}$ such that $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}}) \leq d - r_{\bar{\boldsymbol{E}}} + 1$, we have $\omega(\boldsymbol{y}^\top \boldsymbol{E}) \leq \omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}})$. Thus, this proof is divided into three parts:

(1) the upper-bound of $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}})$: Given that both $\{\mathcal{F}_1^k, \mathcal{F}_2^k, \mathcal{F}_3^k, \boldsymbol{e}^k\}$ and $\{\bar{\mathcal{F}}_1^k, \bar{\mathcal{F}}_2^k, \bar{\mathcal{F}}_3^k, \bar{\boldsymbol{e}}^k\}$ a constitute a decomposition of $\mathcal{X}$. For any $\boldsymbol{y}$, we have

$$\begin{aligned}&[\mathrm{vec}(\mathrm{TR}(\mathcal{F}_1^1, \mathcal{F}_2^1, \mathcal{F}_3^1)), \cdots, \mathrm{vec}(\mathrm{TR}(\mathcal{F}_1^d, \mathcal{F}_2^d, \mathcal{F}_3^d))]\boldsymbol{E}^\top \boldsymbol{y}\\ =&[\mathrm{vec}(\mathrm{TR}(\bar{\mathcal{F}}_1^1, \bar{\mathcal{F}}_2^1, \bar{\mathcal{F}}_3^1)), \cdots, \mathrm{vec}(\mathrm{TR}(\bar{\mathcal{F}}_1^d, \bar{\mathcal{F}}_2^d, \bar{\mathcal{F}}_3^d))]\bar{\boldsymbol{E}}^\top \boldsymbol{y}.\end{aligned} \tag{2}$$

Since $\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3$ are full column rank matrices, they exhibit linear independence for any column vector. By partitioning $\boldsymbol{F}_1$ into submatrices $\boldsymbol{F}_1 = [(\mathcal{F}_1^1)_{(3)}^\top, \cdots, (\mathcal{F}_1^d)_{(3)}^\top] = [\boldsymbol{F}_1^1, \cdots, \boldsymbol{F}_1^d]$, it follows that these submatrices are also linearly independent, i.e., any specific submatrix $\boldsymbol{F}_1^k$ cannot be expressed as a linear combination of the others $\boldsymbol{F}_1^{\neq k}$. This linear independence extends similarly to the submatrices of $\boldsymbol{F}_2$ and $\boldsymbol{F}_3$. Due to $\mathcal{Z}_k = \mathrm{TR}(\mathcal{F}_1^k, \mathcal{F}_2^k, \mathcal{F}_3^k)$, the space spanned by $\mathcal{Z}$, denoted as span($\mathcal{Z}^k$), is generated by the span($\mathcal{F}_1^k$), span($\mathcal{F}_2^k$), and span($\mathcal{F}_3^k$). Therefore, we have span$\left(\mathcal{Z}^k\right) \subseteq$ span$\left(\mathcal{F}_1^k\right)$ +span$\left(\mathcal{F}_2^k\right)$+span$\left(\mathcal{F}_3^k\right)$. And since any corresponding block matrices of $\mathcal{F}_1^k$, $\mathcal{F}_2^k$, and $\mathcal{F}_3^k$ are linearly independent, respectively, then the intersections are $\bigcap_{k=1}^d \mathit{span}\left(\mathcal{F}_1^k\right) = \{0\}$, $\bigcap_{k=1}^d \mathrm{span}\left(\mathcal{F}_2^k\right) = \{0\}$, and $\bigcap_{k=1}^d \mathrm{span}\left(\mathcal{F}_3^k\right) = \{0\}$, respectively. From this, we conclude $\bigcap_{k=1}^d \mathrm{span}\left(\mathcal{Z}^k\right) = \{0\}$. Then, we deduce that $\mathcal{Z}^k$ is linearly independent and $[\mathrm{vec}(\mathcal{Z}_1), \cdots, \mathrm{vec}(\mathcal{Z}^d)]$ is full column rank matrix. Considering Equation (2), above result implies that if $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}}) = 0$, then $\omega(\boldsymbol{y}^\top \boldsymbol{E}) = 0$ is hold, indicating that null($\bar{\boldsymbol{E}}$) $\subseteq$ null($\boldsymbol{E}$). It implies that span($\boldsymbol{E}$) $\subseteq$ span($\bar{\boldsymbol{E}}$) and $r_{\boldsymbol{E}} \leq r_{\bar{\boldsymbol{E}}}$. Furthermore, if $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}}) \leq d - r_{\bar{\boldsymbol{E}}} + 1$, then we derive

$$\omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{E}}\right) \leq d - r_{\bar{\boldsymbol{E}}} + 1 \leq d - r_{\boldsymbol{E}} + 1 \leq d - k'_{\boldsymbol{E}} + 1 \leq k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} - d - 2. \tag{3}$$

(2) the lower-bound of $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}})$: Based on the definition of BTRD, we have

$$\begin{aligned}\boldsymbol{X}_{n_2n_3\times n_4, i_1} &= \left[\boldsymbol{S}_1^1, \cdots, \boldsymbol{S}_1^d\right] \mathrm{blockdiag}\left(\left[(\boldsymbol{F}_1^1)_{i_1 1}, \cdots, (\boldsymbol{F}_1^1)_{i_1 R}\right], \cdots, \left[\left(\boldsymbol{F}_1^d\right)_{i_1 1}, \cdots, \left(\boldsymbol{F}_1^d\right)_{i_1 R}\right]\right) \boldsymbol{E}^\top\\ \boldsymbol{X}_{n_1n_3\times n_4, i_2} &= \left[\boldsymbol{S}_2^1, \cdots, \boldsymbol{S}_2^d\right] \mathrm{blockdiag}\left(\left[(\boldsymbol{F}_2^1)_{i_2 1}, \cdots, (\boldsymbol{F}_2^1)_{i_2 R}\right], \cdots, \left[\left(\boldsymbol{F}_2^d\right)_{i_2 1}, \cdots, \left(\boldsymbol{F}_2^d\right)_{i_2 R}\right]\right) \boldsymbol{E}^\top\\ \boldsymbol{X}_{n_1n_2\times n_4, i_3} &= \left[\boldsymbol{S}_3^1, \cdots, \boldsymbol{S}_3^d\right] \mathrm{blockdiag}\left(\left[(\boldsymbol{F}_3^1)_{i_3 1}, \cdots, (\boldsymbol{F}_3^1)_{i_3 R}\right], \cdots, \left[\left(\boldsymbol{F}_3^d\right)_{i_3 1}, \cdots, \left(\boldsymbol{F}_3^d\right)_{i_3 R}\right]\right) \boldsymbol{E}^\top\\ \boldsymbol{X}_{n_1n_2\times n_3, i_4} &= \left[\boldsymbol{S}_3^1, \cdots, \boldsymbol{S}_3^d\right] \mathrm{blockdiag}\left(e_{i_4 1}\boldsymbol{I}, \cdots, e_{i_4 d}\boldsymbol{I}\right) \left[\boldsymbol{F}_3^1, \cdots, \boldsymbol{F}_3^d\right]^\top,\end{aligned} \tag{4}$$

Sheng Liu, Xi-Le Zhao, and Hao Zhang are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (email: liusheng16@163.com; xlzhao122003@163.com; aaronzhangfy@163.com;), Corresponding author: Xi-Le Zhao.

where $\boldsymbol{S}_i^k \triangleq (\boldsymbol{S}^{\neq i})^k$ for $i \in [3]$ and $k \in [d]$, and $(\boldsymbol{S}^{\neq i})^k$ satisfies $\mathrm{TR}(\mathcal{F}_1^k, \mathcal{F}_2^k, \mathcal{F}_3^k)_{(i)} = \boldsymbol{F}_i^k((\boldsymbol{S}^{\neq i})^k)^\top$. Consider the linear combination $\sum_{k=1}^{n_4} \boldsymbol{y}_{i_4} \boldsymbol{X}_{n_1 n_2 \times n_3, i_4}$, it is given by

$$\begin{aligned}&\left[\boldsymbol{S}_3^1, \cdots, \boldsymbol{S}_3^d\right] \text{blockdiag}\left(\boldsymbol{y}^\top \boldsymbol{e}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \boldsymbol{e}_d \boldsymbol{I}\right) \left[\boldsymbol{F}_3^1, \cdots, \boldsymbol{F}_3^d\right]^\top \\ =&\left[\bar{\boldsymbol{S}}_3^1, \cdots, \bar{\boldsymbol{S}}_3^d\right] \text{blockdiag}\left(\boldsymbol{y}^\top \bar{\boldsymbol{e}}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \bar{\boldsymbol{e}}_d \boldsymbol{I}\right) \left[\bar{\boldsymbol{F}}_3^1, \cdots, \bar{\boldsymbol{F}}_3^d\right]^\top.\end{aligned} \tag{5}$$

Let $\boldsymbol{S}_3 \triangleq \left[\boldsymbol{S}_3^1, \cdots, \boldsymbol{S}_3^d\right]$, we have

$$\begin{aligned}R\omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{E}}\right) = r_{\text{blockdiag}(\boldsymbol{y}^\top \bar{\boldsymbol{e}}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \bar{\boldsymbol{e}}_d \boldsymbol{I})} &\geq r_{\bar{\boldsymbol{S}}_3 \text{blockdiag}(\boldsymbol{y}^\top \bar{\boldsymbol{e}}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \bar{\boldsymbol{e}}_d \boldsymbol{I}) \bar{\boldsymbol{F}}_3} \\ &= r_{\boldsymbol{S}_3 \text{blockdiag}(\boldsymbol{y}^\top \boldsymbol{e}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \boldsymbol{e}_d \boldsymbol{I}) \boldsymbol{F}_3}.\end{aligned} \tag{6}$$

Let $\gamma = \omega\left(\boldsymbol{y}^\top \boldsymbol{E}\right)$, $\hat{\boldsymbol{S}}_3$ and $\hat{\boldsymbol{F}}_3$ consist of the submatrices of $\boldsymbol{S}_3$ and $\boldsymbol{F}_3$, respectively, corresponding to the nonzero elements of $\boldsymbol{y}^\top \boldsymbol{E}$. Then $\hat{\boldsymbol{S}}_3$ and $\hat{\boldsymbol{F}}_3$ both have γR columns. Let be the $(\gamma \times 1)$ vector containing the nonzero elements of $\boldsymbol{y}^\top \boldsymbol{E}$ such that

$$\boldsymbol{S}_3 \,\text{blockdiag}\left(\boldsymbol{y}^\top \boldsymbol{e}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \boldsymbol{e}_d \boldsymbol{I}\right) \boldsymbol{F}_3^\top = \hat{\boldsymbol{S}}_3 \,\text{blockdiag}\left(\boldsymbol{u}_1 \boldsymbol{I}, \cdots, \boldsymbol{u}_\gamma \boldsymbol{I}\right) \hat{\boldsymbol{F}}_3^\top. \tag{7}$$

And then, we have

$$\begin{aligned}&r_{\boldsymbol{S}_3 blockdiag(\boldsymbol{y}^\top \boldsymbol{e}_1 \boldsymbol{I}, \cdots, \boldsymbol{y}^\top \boldsymbol{e}_d \boldsymbol{I}) \boldsymbol{F}_3} = r_{\hat{\boldsymbol{S}}_3 blockdiag(\boldsymbol{u}_1 \boldsymbol{I}, \cdots, \boldsymbol{u}_\gamma \boldsymbol{I}) \hat{\boldsymbol{F}}_3} \\ &\geq r_{\hat{\boldsymbol{S}}_3} + r_{blockdiag(\boldsymbol{u}_1 \boldsymbol{I}, \cdots, \boldsymbol{u}_\gamma \boldsymbol{I}) \hat{\boldsymbol{F}}_3} - \gamma d \geq r_{\hat{\boldsymbol{S}}_3} + r_{\hat{\boldsymbol{F}}_3} - \gamma R\end{aligned} \tag{8}$$

From the definition of k'-rank, we have

$$r_{\hat{\boldsymbol{S}}_3} \geq R \min(\gamma, k'_{\boldsymbol{S}_3}), r_{\hat{\boldsymbol{F}}_3} \geq R \min(\gamma, k'_{\boldsymbol{F}_3}) \tag{9}$$

Combination of Equation (6), Equation (8), and Equation (9), we can deduce the lower bound of the lower-bound of $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}})$ as follows:

$$\min(\gamma, k'_{\boldsymbol{S}_3}) + \min(\gamma, k'_{\boldsymbol{F}_3}) - \gamma \leq \omega(\boldsymbol{y}^\top \bar{\boldsymbol{E}}). \tag{10}$$

(3) Combination of the two bounds: Combination of Equation (10) and Equation (3), we have

$$\min(\gamma, k'_{\boldsymbol{S}_3}) + \min(\gamma, k'_{\boldsymbol{F}_3}) - \gamma \leq \omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{E}}\right) \leq k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} - d - 2. \tag{11}$$

Let $k'_{\boldsymbol{S}_3} = \varsigma$. Based on the definition of k'-rank, there exists $\{\boldsymbol{S}_3^{p_1}, \cdots, \boldsymbol{S}_3^{p_\varsigma}\}$ form $\{\boldsymbol{S}_3^1, \cdots, \boldsymbol{S}_3^d\}$ and $\mu_1 \neq 0, \mu_2 \neq, \cdots, \mu_\varsigma \neq 0$, such that

$$\begin{aligned}&\mu_1 \boldsymbol{S}_3^{p_1} + \mu_2 \boldsymbol{S}_3^{p_2} + \cdots + \mu_\varsigma \boldsymbol{S}_3^{p_\varsigma} = 0 \\ \Leftrightarrow &\mu_1 \mathcal{S}_3^{p_1} + \mu_2 \mathcal{S}_3^{p_2} + \cdots + \mu_\varsigma \mathcal{S}_3^{p_\varsigma} = 0 \\ \Leftrightarrow &\mu_1 \left(\mathcal{S}_3^{p_1}\right)_{(3)} + \mu_2 \left(\mathcal{S}_3^{p_2}\right)_{(3)} + \cdots + \mu_\varsigma \left(\mathcal{S}_3^{p_\varsigma}\right)_{(3)} = 0\end{aligned} \tag{12}$$

Based on definition of subchain tensor $\mathcal{S}_3^k$, we have

$$\begin{aligned}&\mu_1 \left(\mathcal{S}_3^{p_1}\right)_{(3)} + \mu_2 \left(\mathcal{S}_3^{p_2}\right)_{(3)} + \cdots + \mu_\varsigma \left(\mathcal{S}_3^{p_\varsigma}\right)_{(3)} = 0 \\ \Leftrightarrow &\left[\left(\mathcal{F}_1^{p_1}\right)_{(3)}^\top, \cdots, \left(\mathcal{F}_1^{p_\varsigma}\right)_{(3)}^\top\right] block\text{diag}\left(\mu_1 \boldsymbol{I}, \cdots, \mu_\varsigma \boldsymbol{I}\right) \left[\left(\mathcal{F}_2^{p_1}\right)_{(1)}^\top, \cdots, \left(\mathcal{F}_2^{p_\varsigma}\right)_{(1)}^\top\right] = 0 \\ \Leftrightarrow &\tilde{\boldsymbol{F}}_1 block\text{diag}\left(\mu_1 \boldsymbol{I}, \cdots, \mu_\varsigma \boldsymbol{I}\right) \tilde{\boldsymbol{F}}_2 = 0,\end{aligned} \tag{13}$$

where $\tilde{\boldsymbol{F}}_1 = \left[\left(\mathcal{F}_1^{p_1}\right)_{(3)}^\top, \cdots, \left(\mathcal{F}_1^{p_\varsigma}\right)_{(3)}^\top\right]$ and $\tilde{\boldsymbol{F}}_2 = \left[\left(\mathcal{F}_2^{p_1}\right)_{(1)}^\top, \cdots, \left(\mathcal{F}_2^{p_\varsigma}\right)_{(1)}^\top\right]$. And then, we have

$$r_{\tilde{\boldsymbol{F}}_1 block\text{diag}(\mu_1 \boldsymbol{I}, \cdots, \mu_\varsigma \boldsymbol{I}) \tilde{\boldsymbol{F}}_2} \geq r_{\tilde{\boldsymbol{F}}_1} + r_{block\text{diag}(\mu_1 \boldsymbol{I}, \cdots, \mu_\varsigma \boldsymbol{I}) \tilde{\boldsymbol{F}}_2} - \varsigma d \geq r_{\tilde{\boldsymbol{F}}_1} + r_{\tilde{\boldsymbol{F}}_2} - \varsigma R \tag{14}$$

From the definition of k'-rank, we have

$$r_{\tilde{\boldsymbol{F}}_1} \geq R \min(\varsigma, k'_{\boldsymbol{F}_1}), r_{\tilde{\boldsymbol{F}}_2} \geq R \min(\varsigma, k'_{\boldsymbol{F}_2}), \tag{15}$$

and thus

$$r_{\tilde{\boldsymbol{F}}_1 block\text{diag}(\mu_1 \boldsymbol{I}, \cdots, \mu_\varsigma \boldsymbol{I}) \tilde{\boldsymbol{F}}_2} \geq R \min\left(k'_{\boldsymbol{F}_1}, \varsigma\right) + R \min\left(k'_{\boldsymbol{F}_2}, \varsigma\right) - \varsigma R. \tag{16}$$

We note that $k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} - 1 = \max(k'_{\boldsymbol{F}_1}, k'_{\boldsymbol{F}_2}) + \min(k'_{\boldsymbol{F}_1}, k'_{\boldsymbol{F}_2}) - 1 \geq \max(k'_{\boldsymbol{F}_1}, k'_{\boldsymbol{F}_2})$ due to $k'_{\boldsymbol{F}_1} \geq 1, k'_{\boldsymbol{F}_2} \geq 1$. According to Equation (13) is a zero matrix, we deduce that $r_{\tilde{\boldsymbol{F}}_1 block\text{diag}(\mu_1 \boldsymbol{I}, \cdots, \mu_\varsigma \boldsymbol{I}) \tilde{\boldsymbol{F}}_2} = 0$. Suppose $k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} - 1 \geq \varsigma \geq \max(k'_{\boldsymbol{F}_1}, k'_{\boldsymbol{F}_2})$, we deduce that $k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} - 1 \geq \varsigma \geq k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2}$ is hold via Equation (16), which is impossible. Suppose $k'_{\boldsymbol{F}_1} \geq \varsigma \geq k'_{\boldsymbol{F}_2}$, we deduce $0 \geq k'_{\boldsymbol{F}_2}$ via Equation (16), which is impossible. Suppose $k'_{\boldsymbol{F}_2} \geq \varsigma \geq k'_{\boldsymbol{F}_1}$, we deduce $0 \geq k'_{\boldsymbol{F}_1}$ via Equation (16), which is impossible. Suppose $\min(k'_{\boldsymbol{F}_1}, k'_{\boldsymbol{F}_2}) \geq \varsigma$, we deduce that $0 \geq \varsigma$ via Equation (16), which is impossible. Based on the above analysis, we have

$$k'_{\boldsymbol{S}_3} = \varsigma \geq k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} - 1. \tag{17}$$

Similarly, we analyze Equation (11). Suppose $\gamma > \max(k'_{\boldsymbol{S}_3}, k'_{\boldsymbol{F}_3})$, we can deduce $\gamma \geq d+1$ via Equation (11), which is impossible. Suppose $k'_{\boldsymbol{F}_3} \geq \gamma \geq k'_{\boldsymbol{S}_3}$, we can deduce $k'_{\boldsymbol{F}_3} \geq d+1$, which is impossible. Suppose $k'_{\boldsymbol{S}_3} \geq \gamma \geq k'_{\boldsymbol{F}_3}$, we can deduce $k'_{\boldsymbol{S}_3} \geq k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} - 1 \geq d+1$, which is impossible. Thus, we have $\gamma < \min(k'_{\boldsymbol{S}_3}, k'_{\boldsymbol{F}_3})$, which implies $\omega\left(\boldsymbol{y}^\top \boldsymbol{E}\right) \leq \omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{E}}\right)$. Applied Lemma 1, there holds $\bar{\boldsymbol{E}} = \boldsymbol{E}\Pi_{\boldsymbol{E}}\Lambda_{\boldsymbol{E}}$, $\Pi_{\boldsymbol{E}}$ is a permutation matrix and $\Lambda_{\boldsymbol{E}}$ a nonsingular diagonal matrix. $\square$

Lemma 3. *For $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3\times n_4}$, its BTRD is $\mathcal{X} = \sum_{k=1}^{d} \boldsymbol{e}^k \circ TR\left(\mathcal{F}_1^k, \mathcal{F}_2^k, \mathcal{F}_3^k\right)$ with $R_1 = R_2 = R$ and $R_3 = 1$. Suppose $\boldsymbol{F}_2, \boldsymbol{F}_3$ and $\boldsymbol{E}$ are full column rank matrices satisfy*

$$k'_{\boldsymbol{F}_1} + k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} + k'_{\boldsymbol{E}} + \geq 2d+3, \tag{18}$$

Then, we have an alternative decomposition of $\mathcal{X}$, i.e., $\mathcal{X} = \sum_{k=1}^{d} \bar{\boldsymbol{e}}^k \circ TR\left(\bar{\mathcal{F}}_1^k, \bar{\mathcal{F}}_2^k, \bar{\mathcal{F}}_3^k\right)$, where there holds $\bar{\boldsymbol{F}}_1 = \boldsymbol{F}_1\Pi_{\boldsymbol{F}_1}\Lambda_{\boldsymbol{F}_1}$, $\Pi_{\boldsymbol{F}_1}$ is a permutation matrix and $\Lambda_{\boldsymbol{F}_1}$ a nonsingular diagonal matrix.

Proof. Similar to the proof of Lemma 2, this proof is divided into three parts:

(1) the upper-bound of $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1)$: As $\{\mathcal{F}_1^k, \mathcal{F}_2^k, \mathcal{F}_3^k, \boldsymbol{e}^k\}$ and $\{\bar{\mathcal{F}}_1^k, \bar{\mathcal{F}}_2^k, \bar{\mathcal{F}}_3^k, \bar{\boldsymbol{e}}^k\}$ are both represent a decomposition of $\boldsymbol{X}$. Due to $\boldsymbol{E}$ is full column rank, there are $\{\boldsymbol{e}_{i_41}, \cdots \boldsymbol{e}_{i_4d}\}$ which are different from zero. For any $\boldsymbol{y}$, we have

$$\begin{aligned} &\left[\boldsymbol{S}_1^1, \cdots, \boldsymbol{S}_1^d\right] \text{blockdiag}\left(\boldsymbol{e}_{i_41}\boldsymbol{I}, \cdots, \boldsymbol{e}_{i_4d}\boldsymbol{I}\right) \left[\boldsymbol{F}_1^1, \cdots, \boldsymbol{F}_1^d\right]^\top \boldsymbol{y} \\ =& \left[\bar{\boldsymbol{S}}_1^1, \cdots, \bar{\boldsymbol{S}}_1^d\right] \text{blockdiag}\left(\bar{\boldsymbol{e}}_{i_41}\boldsymbol{I}, \cdots, \bar{\boldsymbol{e}}_{i_4d}\boldsymbol{I}\right) \left[\bar{\boldsymbol{F}}_1^1, \cdots, \bar{\boldsymbol{F}}_1^d\right]^\top \boldsymbol{y}. \end{aligned} \tag{19}$$

Since $\boldsymbol{F}_2$ and $\boldsymbol{F}_3$ are full column rank, We have that $\left[\boldsymbol{S}_1^1, \cdots, \boldsymbol{S}_1^d\right]$ is a column full matrix. And then $\{\boldsymbol{e}_{i_41}, \cdots \boldsymbol{e}_{i_4d}\}$ which is different from zero yields $\text{blockdiag}\left(\boldsymbol{e}_{i_41}\boldsymbol{I}, \cdots, \boldsymbol{e}_{i_4d}\boldsymbol{I}\right)$ with full column rank property. This result implies that if $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1) = 0$, then also $\omega(\boldsymbol{y}^\top \boldsymbol{F}_1) = 0$, which shows null($\bar{\boldsymbol{F}}_1$) $\subseteq$ null($\boldsymbol{F}_1$). It states that $r_{\boldsymbol{F}_1} \leq r_{\bar{\boldsymbol{F}}_1}$. If $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1) \leq d - r_{\bar{\boldsymbol{F}}_1} + 1$, then we have

$$\omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1\right) \leq d - r_{\bar{\boldsymbol{F}}_1} + 1 \leq d - r_{\boldsymbol{F}_1} + 1 \leq d - k'_{\boldsymbol{F}_1} + 1 \leq + k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} + k'_{\boldsymbol{E}} - d - 2. \tag{20}$$

(2) the lower-bound of $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1)$: Consider the linear combination $\sum_{k=1}^{n_1} \boldsymbol{y}_{i_1}\boldsymbol{X}_{n_2n_3\times n_4, i_1}$, it is given by

$$\begin{aligned} &\left[\boldsymbol{S}_1^1, \cdots, \boldsymbol{S}_1^d\right] \text{blockdiag}\left(\left[\left(\boldsymbol{F}_1^1\right)_{i_11}, \cdots, \left(\boldsymbol{F}_1^1\right)_{i_1R}\right], \cdots, \left[\left(\boldsymbol{F}_1^d\right)_{i_11}, \cdots, \left(\boldsymbol{F}_1^d\right)_{i_1R}\right]\right) \boldsymbol{E}^\top \\ =& \left[\bar{\boldsymbol{S}}_1^1, \cdots, \bar{\boldsymbol{S}}_1^d\right] \text{blockdiag}\left(\left[\left(\bar{\boldsymbol{F}}_1^1\right)_{i_11}, \cdots, \left(\bar{\boldsymbol{F}}_1^1\right)_{i_1R}\right], \cdots, \left[\left(\bar{\boldsymbol{F}}_1^d\right)_{i_11}, \cdots, \left(\bar{\boldsymbol{F}}_1^d\right)_{i_1R}\right]\right) \bar{\boldsymbol{E}}^\top \end{aligned} \tag{21}$$

And then, we have

$$\begin{aligned} R\omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1\right) = r_{\text{blockdiag}\left(\boldsymbol{y}^\top \boldsymbol{T}^1 \cdots, \boldsymbol{y}^\top \boldsymbol{T}^d\right)} &\geq r_{\bar{\boldsymbol{S}}_1 \text{blockdiag}\left(\boldsymbol{y}^\top \boldsymbol{T}^1 \cdots, \boldsymbol{y}^\top \boldsymbol{T}^d\right)\boldsymbol{E}^\top} \\ &= r_{\boldsymbol{S}_1 \text{blockdiag}\left(\boldsymbol{y}^\top \boldsymbol{T}^1, \cdots, \boldsymbol{y}^\top \boldsymbol{T}^d\right)\boldsymbol{E}^\top} \end{aligned} \tag{22}$$

where $\boldsymbol{T}^k \triangleq \left[\left(\boldsymbol{F}_1^k\right)_{i_11}, \cdots, \left(\boldsymbol{F}_1^k\right)_{i_1R}\right]$. Let $\gamma = \omega\left(\boldsymbol{y}^\top \boldsymbol{F}_1\right)$, $\hat{\boldsymbol{S}}_1$ and $\hat{\boldsymbol{E}}$ consist of the submatrices of $\boldsymbol{S}_1$ and $\boldsymbol{E}$, respectively, corresponding to the nonzero elements of $\boldsymbol{y}^\top \boldsymbol{F}_1$. Then $\hat{\boldsymbol{S}}_1$ and $\hat{\boldsymbol{E}}$ both have columns.

$$r_{\boldsymbol{S}_1 \text{blockdiag}\left(\boldsymbol{y}^\top \boldsymbol{T}^1, \cdots, \boldsymbol{y}^\top \boldsymbol{T}^d\right)\boldsymbol{E}^\top} \geq r_{\hat{\boldsymbol{S}}_1} + r_{\hat{\boldsymbol{E}}} - \gamma R. \tag{23}$$

From the definition of k'-rank, we have

$$r_{\hat{\boldsymbol{S}}_1} \geq R\min(\gamma, k'_{\boldsymbol{S}_1}), r_{\hat{\boldsymbol{E}}} \geq R\min(\gamma, k'_{\boldsymbol{E}}) \tag{24}$$

Combination of Equation (22), Equation (23), and Equation (24), we can deduce the lower bound of the lower-bound of $\omega(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1)$ as follows:

$$\min(\gamma, k'_{\boldsymbol{S}_1}) + \min(\gamma, k'_{\boldsymbol{E}}) - \gamma \leq \omega(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1). \tag{25}$$

(3) Combination of the two bounds: Combination of Equation (25) and Equation (20), we have

$$\min(\gamma, k'_{\boldsymbol{S}_1}) + \min(\gamma, k'_{\boldsymbol{E}}) - \gamma \leq \omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1\right) \leq k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} + k'_{\boldsymbol{E}} - d - 2. \tag{26}$$

Similarly for the derivation of Equation (17), based on definition of subchain tensor $\mathcal{S}_1^k$, we deduce $k'_{\boldsymbol{S}_1} \geq k'_{\boldsymbol{F}_2} + k'_{\boldsymbol{F}_3} - 1$. Suppose $\gamma > \max(k'_{\boldsymbol{S}_1}, k'_{\boldsymbol{E}})$, we can deduce $\gamma \geq d+1$ via Equation (26), which is impossible. Suppose $k'_{\boldsymbol{E}} \geq \gamma \geq k'_{\boldsymbol{S}_1}$, we can deduce $k'_{\boldsymbol{E}} \geq d+1$, which is impossible. Suppose $k'_{\boldsymbol{S}_1} \geq \gamma \geq k'_{\boldsymbol{E}}$, we can deduce $k'_{\boldsymbol{S}_1} \geq d+1$, which is impossible. Thus, we have $\gamma < \min(k'_{\boldsymbol{S}_1}, k'_{\boldsymbol{E}})$, which implies $\omega\left(\boldsymbol{y}^\top \boldsymbol{F}_1\right) \leq \omega\left(\boldsymbol{y}^\top \bar{\boldsymbol{F}}_1\right)$. Applied Lemma 1, there holds $\bar{\boldsymbol{F}}_1 = \boldsymbol{F}_1\Pi_{\boldsymbol{F}_1}\Lambda_{\boldsymbol{F}_1}$, $\Pi_{\boldsymbol{F}_1}$ is a permutation matrix and $\Lambda_{\boldsymbol{F}_1}$ a nonsingular diagonal matrix. $\square$

Remark 1. *There holds $\bar{\boldsymbol{F}}_2=\boldsymbol{F}_2\Pi_{\boldsymbol{F}_2}\Lambda_{\boldsymbol{F}_2}$ with similar conditions for Lemma 3, where $\Pi_{\boldsymbol{F}_2}$ is a permutation matrix and $\Lambda_{\boldsymbol{F}_2}$ a nonsingular diagonal matrix. And, there holds $\bar{\boldsymbol{F}}_3=\boldsymbol{F}_3\Pi_{\boldsymbol{F}_3}\Lambda_{\boldsymbol{F}_3}$ with similar conditions for Lemma 3, where $\Pi_{\boldsymbol{F}_3}$ is a permutation matrix and $\Lambda_{\boldsymbol{F}_3}$ a nonsingular diagonal matrix. These proofs are similar to Lemma 3 and will not be repeated here.*

Based on the above Lemmas, we accomplish the proof of Theorem 1.

Proof. Based on the definition of BTRD, we have

$$\begin{aligned}\boldsymbol{X}_{(4)}^{\top}=&[\operatorname{vec}(\operatorname{TR}(\mathcal{F}_1^1,\mathcal{F}_2^1,\mathcal{F}_3^1)),\cdots,\operatorname{vec}(\operatorname{TR}(\mathcal{F}_1^d,\mathcal{F}_2^d,\mathcal{F}_3^d))]\boldsymbol{E}^{\top}\\=&[\operatorname{vec}(\operatorname{TR}(\bar{\mathcal{F}}_1^1,\bar{\mathcal{F}}_2^1,\bar{\mathcal{F}}_3^1)),\cdots,\operatorname{vec}(\operatorname{TR}(\bar{\mathcal{F}}_1^d,\bar{\mathcal{F}}_2^d,\bar{\mathcal{F}}_3^d))]\bar{\boldsymbol{E}}^{\top}.\end{aligned}\tag{27}$$

According to lemma 2, we can get $\bar{\boldsymbol{E}}=\boldsymbol{E}\Pi_{\boldsymbol{E}}\Lambda_{\boldsymbol{E}}$. Since $k_E'=d$, $\boldsymbol{E}$ is full column rank, and then we have

$$\begin{aligned}&[\operatorname{vec}(\operatorname{TR}(\mathcal{F}_1^1,\mathcal{F}_2^1,\mathcal{F}_3^1)),\cdots,\operatorname{vec}(\operatorname{TR}(\mathcal{F}_1^d,\mathcal{F}_2^d,\mathcal{F}_3^d))\\=&[\operatorname{vec}(\operatorname{TR}(\bar{\mathcal{F}}_1^1,\bar{\mathcal{F}}_2^1,\bar{\mathcal{F}}_3^1)),\cdots,\operatorname{vec}(\operatorname{TR}(\bar{\mathcal{F}}_1^d,\bar{\mathcal{F}}_2^d,\bar{\mathcal{F}}_3^d))]\Pi_{\boldsymbol{E}}^{\top}\Lambda_{\boldsymbol{E}}^{\top}.\end{aligned}\tag{28}$$

Taking into account that $\operatorname{vec}(\operatorname{TR}(\mathcal{F}_1^k,\mathcal{F}_2^k,\mathcal{F}_3^k))$ is a vector representation of the tensor $\operatorname{TR}(\mathcal{F}_1^k,\mathcal{F}_2^k,\mathcal{F}_3^k)$. This implies that the tensor $\operatorname{TR}(\mathcal{F}_1^k,\mathcal{F}_2^k,\mathcal{F}_3^k)$ are ordered in the same way as the vectors $\bar{\boldsymbol{e}}^k$. Furthermore, according to Lemma 3, if $\bar{\boldsymbol{e}}^k=\lambda\boldsymbol{e}^k$, then we have $\operatorname{TR}(\bar{\mathcal{F}}_1^k,\bar{\mathcal{F}}_2^k,\bar{\mathcal{F}}_3^k)=\lambda^{-1}\operatorname{TR}(\mathcal{F}_1^k,\mathcal{F}_2^k,\mathcal{F}_3^k)$. □

II. The proofs of the Theorem 2 and Theorem 3

A. The proof of the Theorem 2

We first introduce some definitions and lemmas before proving Theorem 2.

Definition 1 (ϵ-net [2])**.** *Let $(\mathscr{L},d)$ be a metric space and $\mathscr{E}\subset\mathscr{L}$. For a given $\epsilon>0$, a subset $\overline{\mathscr{E}}\subseteq\mathscr{E}$ is called an ϵ-net of $\mathscr{E}$ if every point in $\mathscr{E}$ is within a distance of ϵ from some point in $\overline{\mathscr{E}}$, i.e.,*

$$\forall\mathcal{X}\in\mathscr{E},\ \exists\,\mathcal{X}_0\in\overline{\mathscr{E}}:d\left(\mathcal{X},\mathcal{X}_0\right)\leq\epsilon.\tag{29}$$

Definition 2 (ϵ-covering numbers [2])**.** *The smallest possible cardinality of an ϵ-net of $\mathscr{E}$ is called the covering number of $\mathscr{E}$ and is denoted as $\mathcal{N}(\mathscr{E},d,\epsilon)$.*

Lemma 4. *Let $\mathscr{E}_i^k=\{\mathcal{F}_i^k\in\mathbb{R}^{R_{i-1}\times n_i\times R_i}|\|\mathcal{F}_i^k\|_F\leq\zeta_i^k\}$, where $k\in[d]$ and $i\in[N-1]$. The ϵ-net of $\mathscr{E}_i^k$ is denoted as $\overline{\mathscr{E}_i^k}$. If $\overline{\mathcal{F}_i^k}\in\overline{\mathscr{E}_i^k}$, we have $\|\overline{\mathcal{F}_i^k}\|_F\leq\zeta_i^k$.*

Proof. Since the ϵ-net of a set is a subset of that set, we have $\overline{\mathscr{E}_i^k}\subset\mathscr{E}_i^k$. It follows that $\overline{\mathcal{F}_i^k}\in\overline{\mathscr{E}_i^k}\subset\mathscr{E}_i^k$. Therefore, we can further deduce that $\|\overline{\mathcal{F}_i^k}\|_F^2\leq\zeta_i^k$, where $k\in[d]$ and $i\in[N-1]$. □

Lemma 5. *Let $\mathscr{E}_1=\left\{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\cdots\times n_N}|\mathcal{X}=\mathcal{P}_{\Omega^c}\left(\sum_{k=1}^d\boldsymbol{e}^k\circ\operatorname{TR}\left(\mathcal{F}_1^k,\cdots,\mathcal{F}_{N-1}^k\right)\right)+\mathcal{P}_\Omega\left(\hat{\mathcal{X}}\right)\right\}$ and $\mathscr{E}_2=\left\{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\cdots\times n_N}|\mathcal{X}=\sum_{k=1}^d\boldsymbol{e}^k\circ\operatorname{TR}\left(\mathcal{F}_1^k,\cdots,\mathcal{F}_{N-1}^k\right)\right\}$, where $\hat{\mathcal{X}}$ is is a constant tensor, $\|\mathcal{F}_i^k\|_F\leq\zeta_i^k$, $k\in[d]$ and $i\in[N-1]$, and $\|\boldsymbol{E}\|_F\leq\eta$ with $\boldsymbol{E}=[\boldsymbol{e}_1,\cdots,\boldsymbol{e}_d]\in\mathbb{R}^{n_N\times d}$. Then, we have $(\mathscr{E}_1,\|\cdot\|_F,\epsilon)\leq(\mathscr{E}_2,\|\cdot\|_F,\epsilon)$.*

Proof. According to Definition 1 and Definition 2, we have

$$\forall\mathcal{Q}_2\in\mathscr{E}_2,\exists\mathcal{Q}_2'\in\overline{\mathscr{E}_2}:\|\mathcal{Q}_2-\mathcal{Q}_2'\|_F\leq\epsilon.\tag{30}$$

Furthermore, based on the relationship between $\mathscr{E}_1$ and $\mathscr{E}_2$, we obtain

$$\forall\mathcal{Q}_1\in\mathscr{E}_1,\exists\mathcal{Q}_2\in\mathscr{E}_2:\mathcal{Q}_1=\mathcal{P}_{\Omega^C}\left(\mathcal{Q}_2\right)+\mathcal{P}_\Omega\left(\hat{\mathcal{X}}\right)\tag{31}$$

and

$$\forall\mathcal{Q}_2\in\mathscr{E}_1,\exists\mathcal{Q}_1\in\mathscr{E}_2:\mathcal{Q}_1=\mathcal{P}_{\Omega^C}\left(\mathcal{Q}_2\right)+\mathcal{P}_\Omega\left(\hat{\mathcal{X}}\right).\tag{32}$$

Let $\overline{\mathscr{E}_2}$ be the ϵ-net of $\mathscr{E}_2$ with the smallest possible cardinality. Then, we can obtain a set $\overline{\mathscr{E}}$ whose elements are of the form

$$\mathcal{Q}=\mathcal{P}_{\Omega^C}(\mathcal{Q}_2)+\mathcal{P}_\Omega(\hat{\mathcal{X}}),\forall\mathcal{Q}_2\in\overline{\mathscr{E}_2}.\tag{33}$$

According to the above mapping, the cardinality of the set $\overline{\mathscr{E}}$ is less than or equal to the cardinality of $\overline{\mathscr{E}_2}$. Next, we prove that $\overline{\mathscr{E}}$ is an ϵ-net for the set $\mathscr{E}_1$. For every $\mathcal{Q}_1\in\mathscr{E}_1$, and every $\mathcal{Q}\in\overline{\mathscr{E}}$, we have

$$\|\mathcal{Q}_1-\mathcal{Q}\|_F=\|\mathcal{P}_{\Omega^C}\left(\mathcal{Q}_1-\mathcal{Q}\right)\|_F=\|\mathcal{P}_{\Omega^C}\left(\mathcal{Q}_2-\mathcal{Q}\right)\|_F\leq\|\mathcal{Q}_2-\mathcal{Q}\|_F.\tag{34}$$

Therefore, $\overline{\mathscr{E}}$ is a ϵ-net of $\mathscr{E}_1$. Then, we can get $(\mathscr{E}_1,\|\cdot\|_F,\epsilon)\leq(\mathscr{E}_2,\|\cdot\|_F,\epsilon)$. □

Lemma 6. *Let* $\mathscr{E} = \{\mathcal{X} | \mathcal{X} = \sum_{k=1}^{d} \boldsymbol{e}^k \circ \mathrm{TR}\left(\mathcal{F}_1^k, \cdots, \mathcal{F}_{N-1}^k\right)\}$ *and* $\|\mathcal{F}_i^k\|_F \leq \zeta_i^k$ *, where* $k \in [d]$ *and* $i \in [N-1]$*,* $\mathcal{F}_i^k \in \mathbb{R}^{R_{i-1} \times n_i \times R_i}$ *with* $R_0 = R_{N-1}$*, and* $\|\boldsymbol{E}\|_F \leq \eta, \boldsymbol{E} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d] \in \mathbb{R}^{n_N \times d}$*. Then, the covering numbers of* $\mathscr{E}$ *with respect to the Frobenius norm satisfy*

$$\mathcal{N}(\mathscr{E}, \|\cdot\|_F, \epsilon) \leq \left(\frac{3Nd\eta \prod_{k=1}^{d} \prod_{i=1}^{N-1} \zeta_i^k}{\epsilon}\right)^{d\left(n_N + \sum_{i=1}^{N-1} n_i R_{i-1} R_i\right)}. \tag{35}$$

Proof. Let $\mathscr{E}_i^k := \{\boldsymbol{F}_i^k \in \mathbb{R}^{n_i \times R_{i-1} R_i} : \|\boldsymbol{F}_i^k\| \leq \zeta_i^k\}$. Then, there exists an ϵ-net $\overline{\mathscr{E}_i^k}$ such that

$$\mathcal{N}\left(\mathscr{E}_i^k, \|\cdot\|_F, \epsilon\right) \leq \left(\frac{3\zeta_i^k}{\epsilon}\right)^{n_i R_{i-1} R_i}, \tag{36}$$

where $\|\boldsymbol{F}_i^k - \overline{\boldsymbol{F}_i^k}\|_F \leq \epsilon$. By substituting ϵ with ϵ/γ_i^k, we obtain $\|\boldsymbol{F}_i^k - \overline{\boldsymbol{F}_i^k}\|_F \leq \epsilon/\gamma_i^k$ and $\|\boldsymbol{E}_i^k - \overline{\boldsymbol{E}_i^k}\|_F \leq \epsilon/\gamma_0$. Let $\gamma_0 = Nd \prod_{m=1}^{d} \prod_{n=1}^{N-1} \zeta_n^m$ and $\gamma_i^k = Nd\eta \prod_{m \neq k} \prod_{n \neq i} \zeta_n^m$. Therefore, we have

$$\begin{aligned}
&\|\mathcal{X} - \overline{\mathcal{X}}\|_F = \|\sum_{k=1}^{d} \boldsymbol{e}^k \circ \mathcal{Z}^k - \sum_{k=1}^{d} \overline{\boldsymbol{e}^k} \circ \overline{\mathcal{Z}^k}\|_F \\
\leq& \|\sum_{k=1}^{d} \boldsymbol{e}^k \circ \mathcal{Z}^k - \sum_{k=1}^{d} \overline{\boldsymbol{e}^k} \circ \mathcal{Z}^k + \sum_{k=1}^{d} \overline{\boldsymbol{e}^k} \circ \mathcal{Z}^k - \sum_{k=1}^{d} \overline{\boldsymbol{e}^k} \circ \overline{\mathcal{Z}^k}\|_F \\
\leq& \|\sum_{k=1}^{d} \left(\boldsymbol{e}^k - \overline{\boldsymbol{e}^k}\right) \circ \mathcal{Z}^k\|_F + \|\sum_{k=1}^{d} \overline{\boldsymbol{e}^k} \circ \left(\mathcal{Z}^k - \overline{\mathcal{Z}^k}\right)\|_F \\
\leq& \|\boldsymbol{E} - \overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\mathcal{Z}^k\|_F + \|\overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\mathcal{Z}^k - \overline{\mathcal{Z}^k}\|_F \\
=& \|\boldsymbol{E} - \overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\mathcal{Z}^k\|_F + \|\overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\mathrm{TR}\left(\mathcal{F}_1^k, \cdots, \mathcal{F}_{N-1}^k\right) - \mathrm{TR}\left(\overline{\mathcal{F}_1^k}, \cdots, \overline{\mathcal{F}_{N-1}^k}\right)\|_F \\
\leq& \|\boldsymbol{E} - \overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\mathcal{Z}^k\|_F + \|\overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\boldsymbol{F}_1^k \boldsymbol{S}^{\neq 1} - \overline{\boldsymbol{F}_1^k} \boldsymbol{S}^{\neq 1} + \overline{\boldsymbol{F}_1^k} \boldsymbol{S}^{\neq 1} + \cdots - \overline{\boldsymbol{F}_1^k \boldsymbol{S}^{\neq 1}}\|_F \\
\leq& \|\boldsymbol{E} - \overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \|\mathcal{Z}^k\|_F + \|\overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \left(\|\boldsymbol{F}_1^k - \overline{\boldsymbol{F}_1^k}\|_F \|\boldsymbol{S}^{\neq 1}\|_F + \|\boldsymbol{F}_2^k - \overline{\boldsymbol{F}_2^k}\|_F \|\boldsymbol{S}^{\neq 2}\|_F\right) \\
&+ \cdots + \|\overline{\boldsymbol{E}}\|_F \sum_{k=1}^{d} \left(\|\boldsymbol{F}_{N-2}^k - \overline{\boldsymbol{F}_{N-2}^k}\|_F \|\boldsymbol{S}^{\neq N-2}\|_F + \|\boldsymbol{F}_{N-1}^k - \overline{\boldsymbol{F}_{N-1}^k}\|_F \|\boldsymbol{S}^{\neq N-1}\|_F\right) \\
\leq& \frac{\epsilon}{\gamma_0} \prod_{m=1}^{d} \prod_{n=1}^{N-1} \zeta_n^m + \frac{\epsilon}{\gamma_1^1} \prod_{m \neq 1} \prod_{n \neq 1} \zeta_n^m + \frac{\epsilon}{\gamma_2^1} \prod_{m \neq 1} \prod_{n \neq 2} \zeta_n^m + \cdots + \frac{\epsilon}{\gamma_{N-1}^1} \prod_{m \neq 1} \prod_{n \neq N-1} \zeta_n^m \\
&+ \frac{\epsilon}{\gamma_1^2} \prod_{m \neq 2} \prod_{n \neq 1} \zeta_n^m \cdots + \frac{\epsilon}{\gamma_{N-1}^d} \prod_{m \neq d} \prod_{n \neq N-1} \zeta_n^m \\
\leq& \epsilon.
\end{aligned} \tag{37}$$

Based on the above formulas, we know $\overline{\mathscr{E}}$ is an ϵ-net of $\mathscr{E}$. By applying Lemma 5, we can know that the covering number

$$
\begin{aligned}
&\mathcal{N}\left(\mathscr{E},\|\cdot\|_F,\epsilon\right) \\
\leq& \left(\frac{3Nd\prod_{m=1}^{d}\prod_{n=1}^{N-1}\zeta_n^m}{\epsilon}\right)^{n_N d}\prod_{k=1}^{d}\prod_{i=1}^{N-1}\left(\frac{3\zeta_i^k\gamma_i^k}{\epsilon}\right)^{n_iR_{i-1}R_i} \\
\leq& \left(\frac{3Nd\prod_{m=1}^{d}\prod_{n=1}^{N-1}\zeta_n^m}{\epsilon}\right)^{n_N d}\left(\frac{3Nd\eta\prod_{k=1}^{d}\prod_{i=1}^{N-1}\zeta_i^k}{\epsilon}\right)^{\sum_{k=1}^{d}\sum_{i=1}^{N-1}n_iR_{i-1}R_i} \\
\leq& \left(\frac{3Nd\eta\prod_{k=1}^{d}\prod_{i=1}^{N-1}\zeta_i^k}{\epsilon}\right)^{d\left(n_N+\sum_{i=1}^{N-1}n_iR_{i-1}R_i\right)}.
\end{aligned}
\tag{38}
$$

Therefore Lemma 6 holds. □

Lemma 7 ([3]). *Consider $\mathscr{E}$ be a set defined over tensors of size $n_1\times n_2\cdots\times n_N$. Let $|\mathscr{E}|$ be the ϵ covering number of $\mathscr{E}$ with respect to the Frobenius norm, $P=\prod_{i=1}^{N}$, and $|\Omega|$ be the sampling number. Supposing $\mathcal{X}\in\mathscr{E}$ and $\max\{\|\mathcal{O}\|_\infty,\|\mathcal{X}\|_\infty\}\leq\xi$, then*

$$
\sup_{\mathcal{X}\in\mathscr{E}}\left|\frac{\|\mathcal{O}-\mathcal{X}\|_F}{\sqrt{P}}-\frac{\|\mathcal{P}_\Omega\left(\mathcal{O}-\mathcal{X}\right)\|_F}{|\Omega|}\right|\leq\frac{2\epsilon}{\sqrt{|\Omega|}}+\left(\frac{8\xi^4\log\left(|\mathscr{E}|P\right)}{|\Omega|}\right)^{1/4},
\tag{39}
$$

with probability at least $1-2P^{-1}$, where $\mathcal{P}_\Omega(\cdot)$ is projected operator. See [3] for the detailed proof of Lemma 7.

Based on the above lemmas, we accomplish the proof of Theorem 2.

Proof. Based on Lemma 6 and Lemma 7, we derive the inequality

$$
\frac{\|\mathcal{X}-\hat{\mathcal{X}}\|_F}{\sqrt{P}}\leq\frac{\|\mathcal{P}_\Omega\left(\mathcal{X}-\hat{\mathcal{X}}\right)\|_F}{|\Omega|}+\frac{2\epsilon}{\sqrt{|\Omega|}}+\left(\frac{8\xi^4\log\left(|\mathscr{E}|P\right)}{|\Omega|}\right)^{1/4}.
\tag{40}
$$

Given the assumptions, we know that $\|\mathcal{P}_\Omega\left(\mathcal{X}-\hat{\mathcal{X}}\right)\|_F=0$. By substituting the expression for $\mathscr{E}$ into Equation (35), we can obtain

$$
\frac{\|\mathcal{X}-\hat{\mathcal{X}}\|_F}{\sqrt{P}}\leq\frac{2\epsilon}{\sqrt{\Omega}}+\left(\frac{8\xi^4\left(\log P+\left(d\left(n_N+\sum_{i=1}^{N-1}n_iR_{i-1}R_i\right)\right)\log\left(\frac{3Nd\eta\prod_{k=1}^{d}\prod_{i=1}^{N-1}\zeta_i^k}{\epsilon}\right)\right)}{|\Omega|}\right)^{1/4}.
\tag{41}
$$

Setting $\epsilon=3\xi Nd$, Equation (41) can be rewritten

$$
\frac{\|\mathcal{X}-\hat{\mathcal{X}}\|_F}{\sqrt{P}}\leq\frac{6\xi Nd}{\sqrt{\Omega}}+\left(\frac{8\xi^4\log P+\left(d\left(n_N+\sum_{i=1}^{N-1}n_iR_{i-1}R_i\right)\right)H}{|\Omega|}\right)^{1/4},
\tag{42}
$$

where $H=8\xi^4\log(\xi^{-1}\eta\prod_{k=1}^{d}\prod_{i=1}^{N-1}\zeta_i^k)$. □

B. The proof of the Theorem 3

Similarly to the proof of Theorem 2, we first prove Lemma 8 before proving Theorem 3.

Lemma 8. *Let $\mathscr{E}=\{\mathcal{X}|\mathcal{X}=\sum_{k=1}^{d}\boldsymbol{c}^k\circ\left(\boldsymbol{A^k}\boldsymbol{B}^k\right)\}$ and $\|\boldsymbol{A}^k\|_F\leq\zeta_1^k$, $\|\boldsymbol{B}^k\|_F\leq\zeta_2^k$, where $k\in[d]$, $\boldsymbol{A}\in\mathbb{R}^{n_1\times R}$ and $\boldsymbol{B}\in\mathbb{R}^{R\times n_2}$, and $\|\boldsymbol{C}\|_F\leq\eta,\boldsymbol{C}=[\boldsymbol{c}_1,\cdots,\boldsymbol{c}_d]\in\mathbb{R}^{n_3\cdots n_N\times d}$. Then, the covering numbers of $\mathscr{E}$ with respect to the Frobenius norm satisfy*

$$
\mathcal{N}(\mathscr{E},\|\cdot\|_F,\epsilon)\leq\left(\frac{9d\eta\prod_{k=1}^{d}\zeta_1^k\zeta_2^k}{\epsilon}\right)^{d\left(\prod_{i=3}^{N}n_i+(n_1+n_2)R\right)}.
\tag{43}
$$

Proof. Let $\mathscr{E}_1^k:=\{\boldsymbol{A}^k\in\mathbb{R}^{n_1\times R}:\|\boldsymbol{A}^k\|\leq\zeta_1^k\}$. Then, there exists an ϵ-net $\overline{\mathscr{E}_1^k}$ such that

$$
\mathcal{N}\left(\mathscr{E}_1^k,\|\cdot\|_F,\epsilon\right)\leq\left(\frac{3\zeta_1^k}{\epsilon}\right)^{n_1R},
\tag{44}
$$

where $\|\boldsymbol{A}^k - \overline{\boldsymbol{A}^k}\|_F \leq \epsilon$. Similarly, let $\mathscr{E}_2^k := \{\boldsymbol{B}^k \in \mathbb{R}^{R\times n_2} : \|\boldsymbol{B}^k\| \leq \zeta_2^k\}$. Then, there exists an ϵ-net $\overline{\mathscr{E}_2^k}$ such that

$$\mathcal{N}\left(\mathscr{E}_2^k, \|\cdot\|_F, \epsilon\right) \leq \left(\frac{3\zeta_2^k}{\epsilon}\right)^{n_2 R}, \tag{45}$$

where $\|\boldsymbol{B}^k - \overline{\boldsymbol{B}^k}\|_F \leq \epsilon$. By substituting ϵ with ϵ/γ_i^k, we obtain $\|\boldsymbol{A}^k - \overline{\boldsymbol{A}^k}\|_F \leq \epsilon/\gamma_1^k$, $\|\boldsymbol{B}^k - \overline{\boldsymbol{B}^k}\|_F \leq \epsilon/\gamma_2^k$, and $\|\boldsymbol{C} - \overline{\boldsymbol{C}}\|_F \leq \epsilon/\gamma_0$. Let $\gamma_0 = 3d\prod_{m=1}^d \zeta_1^m\zeta_2^m$, $\gamma_1^k = 3d\eta\prod_{m\neq k}\zeta_2^m$, and $\gamma_2^k = 3d\eta\prod_{m\neq k}\zeta_1^m$. Therefore, we have

$$\begin{aligned}
&\|\mathcal{X} - \overline{\mathcal{X}}\|_F = \|\sum_{k=1}^d \boldsymbol{c}^k \circ \left(\boldsymbol{A}^k\boldsymbol{B}^k\right) - \sum_{k=1}^d \overline{\boldsymbol{c}^k} \circ \left(\overline{\boldsymbol{A}^k\boldsymbol{B}^k}\right)\|_F \\
\leq& \|\sum_{k=1}^d \boldsymbol{c}^k \circ \left(\boldsymbol{A}^k\boldsymbol{B}^k\right) - \sum_{k=1}^d \overline{\boldsymbol{c}^k} \circ \left(\boldsymbol{A}^k\boldsymbol{B}^k\right) + \sum_{k=1}^d \overline{\boldsymbol{c}^k} \circ \left(\boldsymbol{A}^k\boldsymbol{B}^k\right) - \sum_{k=1}^d \overline{\boldsymbol{c}^k} \circ \left(\overline{\boldsymbol{A}^k\boldsymbol{B}^k}\right)\|_F \\
\leq& \|\sum_{k=1}^d \left(\boldsymbol{c}^k - \overline{\boldsymbol{c}^k}\right) \circ \left(\boldsymbol{A}^k\boldsymbol{B}^k\right)\|_F + \|\sum_{k=1}^d \overline{\boldsymbol{c}^k} \circ \left(\boldsymbol{A}^k\boldsymbol{B}^k - \overline{\boldsymbol{A}^k\boldsymbol{B}^k}\right)\|_F \\
\leq& \|\boldsymbol{C}\|_F\|\sum_{k=1}^d \boldsymbol{A}^k\boldsymbol{B}^k\|_F + \|\overline{\boldsymbol{C}}\|_F\|\sum_{k=1}^d \left(\boldsymbol{A}^k\boldsymbol{B}^k - \overline{\boldsymbol{A}^k\boldsymbol{B}^k}\right)\|_F \\
\leq& \|\boldsymbol{C}\|_F\|\sum_{k=1}^d \boldsymbol{A}^k\boldsymbol{B}^k\|_F + \|\overline{\boldsymbol{C}}\|_F\|\sum_{k=1}^d \left(\boldsymbol{A}^k - \overline{\boldsymbol{A}^k}\right)\boldsymbol{B}^k\|_F + \|\overline{\boldsymbol{C}}\|_F\|\sum_{k=1}^d \overline{\boldsymbol{A}^k}\left(\boldsymbol{B}^k - \overline{\boldsymbol{B}^k}\right)\|_F \\
\leq& \frac{\epsilon}{\gamma_0}\prod_{m=1}^d \zeta_1^m\zeta_2^m + \frac{\epsilon\eta}{\gamma_1}\prod_{m\neq k}^d \zeta_2^m + \frac{\epsilon\eta}{\gamma_2}\prod_{k=1}^d \zeta_1^m \\
\leq& \epsilon.
\end{aligned}$$

By applying Lemma 5, we can know that the covering number

$$\begin{aligned}
&\mathcal{N}\left(\mathscr{E}, \|\cdot\|_F, \epsilon\right) \\
\leq& \left(\frac{3\cdot 3d\prod_{k=1}^d \zeta_1^k\zeta_2^k}{\epsilon}\right)^{d\prod_{i=3}^N n_i} \prod_{k=1}^d \left(\frac{3\zeta_1^k\gamma_1^k}{\epsilon}\right)^{n_1 R} \prod_{k=1}^d \left(\frac{3\zeta_2^k\gamma_2^k}{\epsilon}\right)^{n_2 R} \\
\leq& \left(\frac{9d\eta\prod_{k=1}^d \zeta_1^k\zeta_2^k}{\epsilon}\right)^{d\left(\prod_{i=3}^N n_i + (n_1+n_2)R\right)}
\end{aligned} \tag{46}$$

□

We accomplish the proof of Theorem 3 based on the this Lemma 8.

Proof. Based on Lemma 8 and Lemma 7, we derive the inequality

$$\frac{\|\mathcal{X} - \hat{\mathcal{X}}\|_F}{\sqrt{P}} \leq \frac{\|\mathcal{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}}\right)\|_F}{|\Omega|} + \frac{2\epsilon}{\sqrt{|\Omega|}} + \left(\frac{8\xi^4\log\left(|\mathscr{E}|P\right)}{|\Omega|}\right)^{1/4}. \tag{47}$$

Given the assumptions, we know that $\|\mathcal{P}_\Omega\left(\mathcal{X} - \hat{\mathcal{X}}\right)\|_F = 0$. By substituting the expression for $\mathscr{E}$ into Equation (43), we can obtain

$$\frac{\|\mathcal{X} - \hat{\mathcal{X}}\|_F}{\sqrt{P}} \leq \frac{2\epsilon}{\sqrt{\Omega}} + \left(\frac{8\xi^4\left(\log P + \left(d\left(\prod_{i=3}^N n_i + (n_1+n_2)R\right)\right)\log\left(\frac{9d\eta\prod_{k=1}^d \zeta_1^k\zeta_2^k}{\epsilon}\right)\right)}{|\Omega|}\right)^{1/4}. \tag{48}$$

Setting $\epsilon = 9\xi d$, Equation (48) can be rewritten

$$\frac{\|\mathcal{X} - \hat{\mathcal{X}}\|_F}{\sqrt{P}} \leq \frac{18\xi d}{\sqrt{\Omega}} + \left(\frac{8\xi^4\log P + \left(d\left(\prod_{i=3}^N n_i + (n_1+n_2)R\right)\right)H}{|\Omega|}\right)^{1/4}, \tag{49}$$

where $H = 8\xi^4\log(\xi^{-1}\eta\prod_{k=1}^d \zeta_1^k\zeta_2^k)$. □

III. The proof of Theorem 4

We first introduce some lemmas before proving Theorem 4.

Lemma 9 (Sufficient decrease lemma [4].). *For any $\rho > 0$, the sequence $(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t)_{t\in\mathbb{N}}$ be generated by Equation (13) (in the main document) satisfies the following formulae:*

$$\begin{cases} \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right) + \frac{\rho}{2}\|\mathcal{X} - \mathcal{X}^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right) + \frac{\rho}{2}\|\mathcal{F}_i^k - \left(\mathcal{F}_i^k\right)^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right) + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{Z}^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^t, \boldsymbol{E}^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^{t+1}\right) + \frac{\rho}{2}\|\boldsymbol{E} - \boldsymbol{E}^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right). \end{cases} \tag{50}$$

Proof. Let $\mathcal{X}^{t+1}, (\mathcal{F}_i^k)^{t+1}, \mathcal{Z}^{t+1}$, and $\boldsymbol{E}^{t+1}$ are minimizers of $\mathcal{L}\left(\mathcal{X}, \mathcal{F}_i^k, \mathcal{Z}, \boldsymbol{E}\right)$. Then, we have

$$\begin{cases} \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right) + \frac{\rho}{2}\|\mathcal{X} - \mathcal{X}^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right) + \frac{\rho}{2}\|\mathcal{F}_i^k - \left(\mathcal{F}_i^k\right)^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right) + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{Z}^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^t, \boldsymbol{E}^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^{t+1}\right) + \frac{\rho}{2}\|\boldsymbol{E} - \boldsymbol{E}^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right). \end{cases} \tag{51}$$

Therefore, the proof of Lemma 9 is completed. □

Lemma 10 (Boundedness). *The sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ be generated by Equation (13) (in the main document) is bounded.*

Proof. According to Lemma 9, it follows that

$$\begin{aligned} \|\left(\mathcal{F}_i^k\right)^t\|_F^2 \leq \sum_{k=1}^{d}\sum_{i=1}^{N-1} \|\left(\mathcal{F}_i^k\right)^t\|_F^2 \leq \mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right) \\ \leq \mathcal{L}\left(\mathcal{X}^{t-1}, \left(\mathcal{F}_i^k\right)^{t-1}, \mathcal{Z}^{t-1}, \boldsymbol{E}^{t-1}\right) \leq \cdots \leq \mathcal{L}\left(\mathcal{X}^0, \left(\mathcal{F}_i^k\right)^0, \mathcal{Z}^0, \boldsymbol{E}^0\right). \end{aligned} \tag{52}$$

Since $\mathcal{L}\left(\mathcal{X}^0, \left(\mathcal{F}_i^k\right)^0, \mathcal{Z}^0, \boldsymbol{E}^0\right)$ is is a constant, $\left(\mathcal{F}_i^k\right)^t$ is bound. By applying the triangle inequality, we obtain

$$\begin{aligned} \|\left(\mathcal{Z}^k\right)^t\|_F^2 - \sum_{i=1}^{N-1} \|\left(\mathcal{F}_i^k\right)^t\|_F^2 \leq \|\left(\mathcal{Z}^k\right)^t\|_F^2 - \mathrm{TR}\left(\left(\mathcal{F}_1^k\right)^t, \cdots, \left(\mathcal{F}_{N-1}^k\right)^t\right) \\ \leq \|\left(\mathcal{Z}^k\right)^t - \mathrm{TR}\left(\left(\mathcal{F}_1^k\right)^t, \cdots, \left(\mathcal{F}_{N-1}^k\right)^t\right)\|_F^2. \end{aligned} \tag{53}$$

This implies

$$\|\mathcal{Z}^t\|_F^2 \leq \sum_{k=1}^{d} \|\left(\mathcal{Z}^k\right)^t\|_F^2 \leq \sum_{k=1}^{d}\left(\sum_{i=1}^{N-1} \|\left(\mathcal{F}_i^k\right)^t\|_F^2 + \|\left(\mathcal{Z}^k\right)^t - \mathrm{TR}\left(\left(\mathcal{F}_1^k\right)^t, \cdots, \left(\mathcal{F}_{N-1}^k\right)^t\right)\|_F^2\right). \tag{54}$$

Suppose $\|\left(\mathcal{Z}^k\right)^t - \mathrm{TR}\left(\left(\mathcal{F}_1^k\right)^t, \cdots, \left(\mathcal{F}_{N-1}^k\right)^t\right)\|_F^2 \to +\infty$, we have

$$\mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right) \to +\infty. \tag{55}$$

This result contradicts Lemma 9, which emphasizes $\mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)$ is bound. Consequently, $\mathcal{Z}^t$ must also be bounded. Based on $\|\boldsymbol{E}(:,k)\|_F^2 = 1$, we can have $\|\boldsymbol{E}^t\|_F^2 \leq d$. This analysis implies that $\boldsymbol{E}^t$ is bound. By applying the triangle inequality, we obtain

$$\|\mathcal{X}^t\|_F^2 - \|\boldsymbol{E}^t\|\|\mathcal{Z}^t\|_F^2 \leq \|\mathcal{X}^t\|_F^2 - \|\sum_{k=1}^{d}\left(\boldsymbol{e}^k\right)^t \circ \left(\mathcal{Z}^k\right)^t\|_F^2 \leq \|\mathcal{X}^t - \sum_{k=1}^{d}\left(\boldsymbol{e}^k\right)^t \circ \left(\mathcal{Z}^k\right)^t\|_F^2,$$

which implies

$$\|\mathcal{X}^t\|_F^2 \leq \|\boldsymbol{E}^t\|\|\mathcal{Z}^t\|_F^2 + \|\mathcal{X}^t - \sum_{k=1}^{d}\left(\boldsymbol{e}^k\right)^t \circ \left(\mathcal{Z}^k\right)^t\|_F^2. \tag{56}$$

Suppose $\|\mathcal{X}^t - \sum_{k=1}^{d}\left(\boldsymbol{e}^k\right)^t \circ \left(\mathcal{Z}^k\right)^t\|_F^2 \to +\infty$, we have

$$\mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right) \to +\infty. \tag{57}$$

This result contradicts Lemma 9, which emphasizes $\mathcal{L}\left(\mathcal{X}^t, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)$ is bound. Thus, $\mathcal{X}^t$ must also be bounded.

In summary, the sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ be generated by **??** is bounded. □

Lemma 11 (Relative error lemma [4].). *For any $\rho > 0$, there exist $\boldsymbol{Q}_w^{t+1}(w = 1, \cdots, 4)$ such that the sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ be generated by Equation (13) (in the main document) is bounded, and satisfies the following formulae:*

$$\begin{cases} \|\boldsymbol{Q}_1^{t+1} + \nabla_{\mathcal{X}}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)\|_F^2 \leq \rho\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_F^2, \\ \|\boldsymbol{Q}_2^{t+1} + \nabla_{\mathcal{F}_i^k}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^t, \boldsymbol{E}^t\right)\|_F^2 \leq \rho\|\left(\mathcal{F}_i^k\right)^{t+1} - \left(\mathcal{F}_i^k\right)^t\|_F^2, \\ \|\boldsymbol{Q}_3^{t+1} + \nabla_{\mathcal{Z}}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right)\|_F^2 \leq \rho\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\|_F^2, \\ \|\boldsymbol{Q}_4^{t+1} + \nabla_{\boldsymbol{E}}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^{t+1}\right)\|_F^2 \leq \rho\|\boldsymbol{E}^{t+1} - \boldsymbol{E}^t\|_F^2, \end{cases} \tag{58}$$

where $\boldsymbol{Q}_1^{t+1} \in \partial\Phi(\mathcal{X})$, $\boldsymbol{Q}_2^{t+1} \in \partial\mathcal{L}_1(\mathcal{F}_i^k)$, $\boldsymbol{Q}_3^{t+1} \in \partial\mathcal{L}_2(\mathcal{Z})$, and $\boldsymbol{Q}_4^{t+1} \in \partial\Psi(\boldsymbol{E})$, $\mathcal{L}_1 = \frac{1}{2}\sum_{k=1}^d \|\mathcal{Z}^k - \mathrm{TR}\left(\mathcal{F}_1^k, \cdots, \mathcal{F}_{N-1}^k\right)\|_F^2$, and $\mathcal{L}_2 = \frac{\alpha}{2}\|\mathcal{X} - \sum_{k=1}^d \mathbf{e}^k \circ \mathcal{Z}^k\|_F^2$.

Proof. Let $\mathcal{X}^{t+1}, (\mathcal{F}_i^k)^{t+1}, \mathcal{Z}^{t+1}$, and $\boldsymbol{E}^{t+1}$ are the optimal solutions of $\mathcal{L}(\mathcal{X}^{t+1}, (\mathcal{F}_i^k)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^{t+1})$. For each subproblem, we have

$$\begin{cases} 0 \in \partial\Phi\left(\mathcal{X}\right) - \alpha\left(\mathcal{X}^t - \sum_{k=1}^d \left(\boldsymbol{e}^k\right)^t \circ \left(\mathcal{Z}^k\right)^t\right) + \rho\left(\mathcal{X}^{t+1} - \mathcal{X}^t\right), \\ 0 \in \partial\mathcal{L}_1\left(\mathcal{F}_i^k\right) - \beta\left(\mathcal{F}_i^k\right)^t + \rho\left(\left(\mathcal{F}_i^k\right)^{t+1} - \left(\mathcal{F}_i^k\right)^t\right) \\ 0 \in \partial\mathcal{L}_2\left(\mathcal{Z}\right) - \sum_{k=1}^d \left(\left(\mathcal{Z}^k\right)^t - \mathrm{TR}\left(\left(\mathcal{F}_1^k\right)^t, \cdots, \left(\mathcal{F}_{N-1}^k\right)^t\right)\right) + \rho\left(\mathcal{Z}^{t+1} - \mathcal{Z}^t\right), \\ 0 \in \partial\Psi\left(\boldsymbol{E}\right) - \alpha\left(\boldsymbol{E}^t\boldsymbol{Z}^{t+1}\left(\boldsymbol{Z}^{t+1}\right)^\top\right) + \rho\left(\boldsymbol{E}^{t+1} - \boldsymbol{E}^t\right). \end{cases} \tag{59}$$

Then, we define $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$ as

$$\begin{cases} \boldsymbol{Q}_1^{t+1} = \alpha\left(\mathcal{X}^t - \sum_{k=1}^d \left(\boldsymbol{e}^k\right)^t \circ \left(\mathcal{Z}^k\right)^t\right) - \rho\left(\mathcal{X}^{t+1} - \mathcal{X}^t\right), \\ \boldsymbol{Q}_2^{t+1} = \beta\left(\mathcal{F}_i^k\right)^t - \rho\left(\left(\mathcal{F}_i^k\right)^{t+1} - \left(\mathcal{F}_i^k\right)^t\right) \\ \boldsymbol{Q}_3^{t+1} = \sum_{k=1}^d \left(\left(\mathcal{Z}^k\right)^t - \mathrm{TR}\left(\left(\mathcal{F}_1^k\right)^t, \cdots, \left(\mathcal{F}_{N-1}^k\right)^t\right)\right) - \rho\left(\mathcal{Z}^{t+1} - \mathcal{Z}^t\right), \\ \boldsymbol{Q}_4^{t+1} = \alpha\left(\boldsymbol{E}^t\boldsymbol{Z}^{t+1}\left(\boldsymbol{Z}^{t+1}\right)^\top\right) - \rho\left(\boldsymbol{E}^{t+1} - \boldsymbol{E}^t\right) \end{cases} \tag{60}$$

Since the sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ is bounded and $\nabla\mathcal{L}$ is Lipschitz continuous on any bounded set, then there exist $\rho > 0$ satisfy the following formulae:

$$\begin{cases} \|\boldsymbol{Q}_1^{t+1} + \nabla_{\mathcal{X}}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)\|_F^2 \leq \rho\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_F^2, \\ \|\boldsymbol{Q}_2^{t+1} + \nabla_{\mathcal{F}_i^k}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^t, \boldsymbol{E}^t\right)\|_F^2 \leq \rho\|\left(\mathcal{F}_i^k\right)^{t+1} - \left(\mathcal{F}_i^k\right)^t\|_F^2, \\ \|\boldsymbol{Q}_3^{t+1} + \nabla_{\mathcal{Z}}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^t, \mathcal{Z}^{t+1}, \boldsymbol{E}^t\right)\|_F^2 \leq \rho\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\|_F^2, \\ \|\boldsymbol{Q}_4^{t+1} + \nabla_{\boldsymbol{E}}\mathcal{L}\left(\mathcal{X}^{t+1}, \left(\mathcal{F}_i^k\right)^{t+1}, \mathcal{Z}^{t+1}, \boldsymbol{E}^{t+1}\right)\|_F^2 \leq \rho\|\boldsymbol{E}^{t+1} - \boldsymbol{E}^t\|_F^2, \end{cases} \tag{61}$$

The proof of Lemma 11 is completed. □

Based on the above preparations, we provide the proof of Theorem 4 about convergence analysis as follows.

Proof. The proof process of Theorem 4 is to refer to Theorem 6.2 in [4], which the proposed algorithm satisfies the following conditions:

1) The sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ satisfies sufficient decrease condition.
2) The sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ satisfies relative error condition.
3) The sequence $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ is bounded;
4) The $\mathcal{L}\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)$ has the Kurdyka-Łojasiewicz (K-Ł) property [4] at $\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$;
5) The $\mathcal{L}\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)$ satisfies continuity condition;

For condition 1), it follows from the proof of Lemma 9 that the sequence $\left(\mathcal{X}^t, (\mathcal{F}i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ satisfies the sufficient decrease condition.

For condition 2), it follows from the proof of Lemma 11 that the sequence $\left(\mathcal{X}^t, (\mathcal{F}i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ satisfies the relative error condition.

For condition 3), it follows from the proof of Lemma 10 that the sequence $\left(\mathcal{X}^t, (\mathcal{F}i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)_{t\in\mathbb{N}}$ is bound.

For condition 4), we demonstrate that $\mathcal{L}\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)$ satisfy the K-Ł property at each point $\mathcal{L}\left(\mathcal{X}^t, (\mathcal{F}_i^k)^t, \mathcal{Z}^t, \boldsymbol{E}^t\right)$ by proving that $\mathcal{L}$ is a semi-algebraic function. The terms $\mathcal{L}_1, \mathcal{L}_2$, and $\frac{\beta}{2}\sum_{k=1}^d \sum_{i=1}^{N-1} \|\mathcal{F}_i^k\|_F^2$ are the Frobenius norm, and

Frobenius norms are semi-algebraic functions [5]. Since $\Phi(\mathcal{X})$ and $\Psi(\boldsymbol{E})$ are indicator functions with semi-algebraic sets, they are semi-algebraic functions [5]. Thus, the function $\mathcal{L}\left(\mathcal{X}^t,(\mathcal{F}_i^k)^t,\mathcal{Z}^t,\boldsymbol{E}^t\right)$ is a semi-algebraic function. Consequently, the function $\mathcal{L}\left(\mathcal{X}^t,(\mathcal{F}_i^k)^t,\mathcal{Z}^t,\boldsymbol{E}^t\right)$ has the K-Ł property at each $\left(\mathcal{X}^t,(\mathcal{F}_i^k)^t,\mathcal{Z}^t,\boldsymbol{E}^t\right)_{t\in\mathbb{N}}$.

For condition 5), the above four conditions ensure that there exists a subsequence $\left(\mathcal{X}^{t_s},\ (\mathcal{F}_i^k)^{t_s},\mathcal{Z}^{t_s},\boldsymbol{E}^{t_s}\right)_{t_s\in\mathbb{N}}$ such that $\lim_{s\to\infty}\mathcal{L}\left(\mathcal{X}^{t_s},(\mathcal{F}_i^k)^{t_s},\mathcal{Z}^{t_s},\boldsymbol{E}^{t_s}\right)$ exists. The objective function $\mathcal{L}\left(\mathcal{X},(\mathcal{F}_i^k),\mathcal{Z},\boldsymbol{E}\right)$ is composed of the L_2 metric function and the indicator function. The indicator function $\Phi(\mathcal{X})$ satisfies $\Phi(\mathcal{X})=0$ when $\mathcal{X}_\Omega=\mathcal{O}_\Omega$ and $\Phi(\mathcal{X})=+\infty$ otherwise. $\mathcal{X}$-based update rule, it satisfies $\mathcal{X}_\Omega^*=\mathcal{O}_\Omega$. Thus, we have $\lim_{s\to\infty}\Phi(\mathcal{X}^{t_s})=\Phi(\mathcal{X}^*)=0$. Similarly, $\boldsymbol{E}$-based update rule, it satisfies $\|\boldsymbol{E}^*(:,k)\|_F^2=1$. Thus, we have $\lim_{s\to\infty}\Psi\left(\boldsymbol{E}^{t_s}\right)=\Psi(\boldsymbol{E}^*)=0$. Since all components of $\mathcal{L}\left(\mathcal{X},(\mathcal{F}_i^k),\mathcal{Z},\boldsymbol{E}\right)$ are continuous at the limit point, it follows that

$$\lim_{s\to\infty}\mathcal{L}\left(\mathcal{X}^{t_s},(\mathcal{F}_i^k)^{t_s},\mathcal{Z}^{t_s},\boldsymbol{E}^{t_s}\right)=\lim_{s\to\infty}\mathcal{L}\left(\mathcal{X}^*,(\mathcal{F}_i^k)^*,\mathcal{Z}^*,\boldsymbol{E}^*\right). \tag{62}$$

$\square$

References

[1] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, 2008.

[2] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47.

[3] J. Fan, "Multi-mode deep matrix and tensor factorization," in *Proceedings of the International Conference on Learning Representations*, 2022.

[4] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods," *Mathematical Programming*, vol. 137, no. 1, pp. 91–129, 2013.

[5] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1, pp. 459–494, 2014.