Supplementary Materials of "Revisiting High-Order Tensor Singular Value Decomposition from Basic Element Perspective"

Sheng Liu, Xi-Le Zhao*, Jinsong Leng, Ben-Zheng Li, Jing-Hua Yang, and Xinyu Chen

In this supplementary document, we provide more experimental results and theoretical proofs. To improve readability, we first give some necessary preliminary knowledge and lemmas. Table I lists some notations and explanations.

TABLE I
NOTATIONS AND EXPLANATIONS.

Notations	Explanations
$[d]$	The set of the first d natural numbers.
x, $\boldsymbol{x}$, $\boldsymbol{X}$, $\mathcal{X}$	Scalar, vector, matrix, tensor.
$\boldsymbol{X}^\top$	The transposition of $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$.
$\sigma_i(\boldsymbol{X})$	The i-th singular value of $\boldsymbol{X}$.
$\\|\boldsymbol{X}\\|_* = \sum \sigma_i(\boldsymbol{X})$	The nuclear norm of $\boldsymbol{X}$.
$\langle \mathcal{X}, \mathcal{Y} \rangle$	The inner product between $\mathcal{X}$ and $\mathcal{Y}$.
$\\|\mathcal{X}\\|_F$	The Frobenius norm of $\mathcal{X}$.
$\boldsymbol{X}^{(i)}$	The i-th frontal slice of third-order tensor $\mathcal{X}$
$\boldsymbol{X}_{(k)} = \text{unfold}_k(\mathcal{X})$	The mode-k matricization of $\mathcal{X}$.
$\mathcal{X} = \text{fold}_k(\boldsymbol{X}_{(k)})$	The inverse operation of $\text{unfold}_k(\mathcal{X})$.

Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be a dth-order tensor, and $\mathcal{X}_{\langle k \rangle} \triangleq \mathscr{P}_{\langle k \rangle}(\mathcal{X}) = \texttt{permute}(\mathcal{X}, [k, \cdots, k+d-1]) \in \mathbb{R}^{n_k \times \cdots \times n_{k+d-1}}$, where $j+d-1 := j$ when $j+d-1 > d$. The corresponding inverse operation as $\mathscr{P}_{\langle k \rangle}^{-1}(\mathcal{X}) = \texttt{ipermute}(\mathcal{X}, [k, \cdots, k+d-1])$, i.e., $\mathcal{X} = \mathscr{P}_{\langle k \rangle}^{-1}(\mathcal{X}_{\langle k \rangle})$.

The mode-k high-dimensional block circulation operation is defined as

$$\text{Hbcirc}(\mathcal{X}, k) = \begin{pmatrix} \boldsymbol{X}_{k+2}^{(1)} & \boldsymbol{X}_{k+d-1}^{(n_{k+d-1})} & \cdots & \boldsymbol{X}_{k+2}^{(2)} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{X}_{k+2}^{(n_{k+2})} & \boldsymbol{X}_{k+2}^{(n_{k+2}-1)} & & \boldsymbol{X}_{k+3}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_{k+d-1}^{(n_{k+d-1})} & \boldsymbol{X}_{k+d-1}^{(n_{k+d-1}-1)} & \cdots & \boldsymbol{X}_{k+2}^{(1)} \end{pmatrix},$$

where $\boldsymbol{X}_k^{(i)} = \mathcal{X}(\underbrace{1, \cdots, 1}_{k-1}, :, :, i, \underbrace{1, \cdots, 1}_{d-k-2}) \in \mathbb{R}^{n_k \times n_{k+1}}$. We define $\boldsymbol{X}_{k+j-1} = \boldsymbol{X}_{k+j-1-d} (j = 1, \cdots, d)$ when $k+j-1 > d$. The mode-k high-dimensional block diagonalization operation and its inverse operation are defined as

$$\text{Hbdiag}(\mathcal{X}, k) = \text{diag}\left(\boldsymbol{X}_{k+1}^{(1)}, \boldsymbol{X}_{k+1}^{(2)}, \cdots, \boldsymbol{X}_{k+d-1}^{(n_{k+d-1})}\right).$$

$$\text{Hbdfold}(\text{Hbdiag}(\mathcal{X}, k), k) = \mathcal{X}.$$

The mode-k high-dimensional block vectorization operation and its inverse operation are defined as

$$\text{Hbvec}(\mathcal{X}, k) = \left(\boldsymbol{X}_{k+2}^{(1)}, \cdots, \boldsymbol{X}_{k+2}^{(n_{k+2})}, \cdots, \boldsymbol{X}_{k+d-1}^{(n_{k+d-1})}\right).$$

$$\text{Hbvfold}(\text{Hbvec}(\mathcal{X}, k), k) = \mathcal{X}.$$

Corresponding author: Xi-Le Zhao.

Sheng Liu, Xi-Le Zhao, Jinsong Leng, and Ben-Zheng Li 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; lengjs@uestc.edu.cn; lbz1604179601@126.com).

Jing-Hua Yang is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: yangjinghua110@126.com).

Xinyu Chen is with the Civil, Geological and Mining Engineering Department, Polytechnique Montreal, Montreal, QC, H3T 1J4, Canada (e-mail: chenxy346@gmail.com).

Definition 1 (Mode-k face-wise product)**.** *For any dth-order tensors* $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_k \times p \times \cdots n_d}$ *and* $\mathcal{Y} \in \mathbb{R}^{n_1 \times \cdots \times p \times n_{k+1} \times \cdots n_d}$*, the mode-*$k$ *face-wise product is denoted as* $\mathcal{X} \triangle_k \mathcal{Y} \in \mathbb{R}^{n_1 \times \cdots \times n \times \cdots \times n_d}$*. For any element-wise, we have*

$$(\mathcal{X} \triangle_k \mathcal{Y})_{i_1 \cdots i_k i_{k+1} \cdots i_d} = \sum_{t=1}^{p} x_{i_1 \cdots i_k t i_{k+2} \cdots i_d} y_{i_1 \cdots i_{k-1} t i_{k+1} \cdots i_d}.$$

Based on Hbdiag *operator, the tensor-matrix product is equivalent to*

$$\mathcal{Z} = \mathcal{X} \triangle_k \mathcal{Y} \Leftrightarrow \mathrm{Hbdiag}(\mathcal{Z}, k) = \mathrm{Hbdiag}(\mathcal{X}, k) \cdot \mathrm{Hbdiag}(\mathcal{Y}, k).$$

Definition 2 (Mode-k tensor-matrix product [1])**.** *For any dth-order tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_k \times \cdots \times n_d}$ *with a matrix* $\boldsymbol{U} \in \mathbb{R}^{n \times n_k}$*, the mode-*$k$ *tensor-matrix product is denoted as* $\mathcal{X} \times_k \boldsymbol{U} \in \mathbb{R}^{n_1 \times \cdots \times n \times \cdots \times n_d}$*. For any element-wise, we have*

$$(\mathcal{X} \times_k \boldsymbol{U})_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_d} = \sum_{i_k=1}^{n_k} x_{i_1 i_2 \cdots i_d} u_{j i_k}.$$

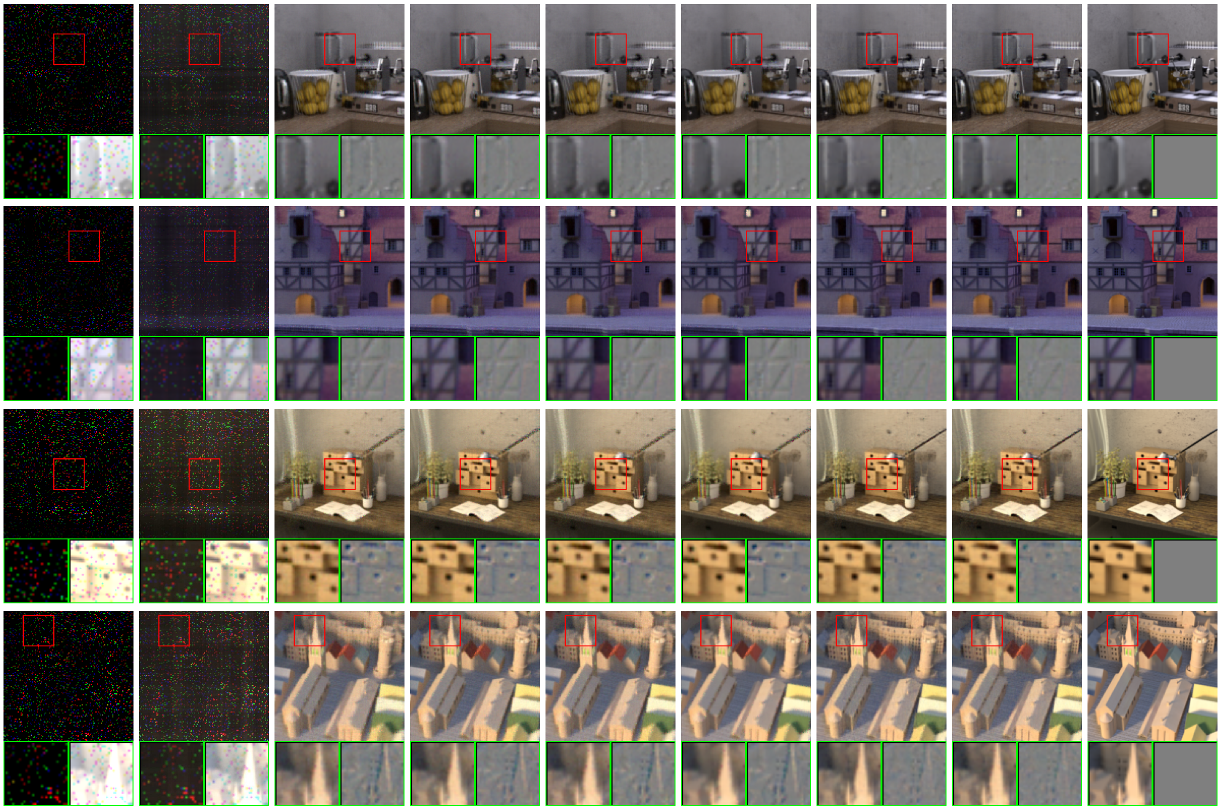
Based on the unfolding operator, the tensor-matrix product is equivalent to

$$\mathcal{Y} = \mathcal{X} \times_k \boldsymbol{U} \Leftrightarrow \boldsymbol{Y}_{(k)} = \boldsymbol{U} \boldsymbol{X}_{(k)}.$$

Appendix A
Experimental results

In this section, we provide more experimental results on color videos and light field images. Fig. 1 presents the recovered results, zoom-in regions, and the corresponding residual images obtained by the HaLRTC, TNN, UTNN, HTNN, WSTNN, METNN, and METNN$^+$ methods at SR=0.2. Observing Fig. 1, it becomes evident that our proposed methods, METNN and METNN$^+$, achieve superior results in both edge profile and fine details compared to the other methods. Fig. 2 shows the visual recovered results by different methods, where SR = 0.05 at *Kitchen*, *Medieval*, *Table*, and *Town*. From Fig. 2, the visual results obtained by the proposed methods outperformed the comparison models on different light field data.

Fig. 1. The selected frame of restoration results by different methods on color videos under SR=0.1. From top to bottom: *Bird*, *Horse*, *Fox*, *Board*, and *Ski*, respectively.



Observed HaLRTC[2] TNN[3] UTNN[4] HTNN[5] WSTNN[6] METNN METNN^{+} Ground truth

Fig. 2. The selected band of restoration results by different methods on light field images under SR=0.05. From top to bottom: *Kitchen*, *Medieval*, *Table*, and *Town*, respectively.

APPENDIX B
THE PROOF OF LEMMA 1

Lemma 1. *Assuming that $\mathcal{G}$, $\mathcal{X}$, and $\mathcal{Y} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ are dth-order tensors. Then, the properties hold as follows.*

(1) $\mathrm{Hbcirc}(\mathcal{X}, k) = \mathrm{Hbcirc}(\mathcal{X}_{\langle k\rangle}, 1)$.
(2) $\mathrm{Hbvec}(\mathcal{X}, k) = \mathrm{Hbvec}(\mathcal{X}_{\langle k\rangle}, 1)$.
(3) $\mathrm{Hbdiag}(\mathcal{X}, k) = \mathrm{Hbdiag}(\mathcal{X}_{\langle k\rangle}, 1)$.
(4) $(\mathcal{X}^{\mathtt{H}_k})_{\langle k\rangle} = (\mathcal{X}_{\langle k\rangle})^{\mathtt{H}_1}$.
(5) $\mathcal{G}_k = \mathcal{X}_k \diamondsuit_k \mathcal{Y}_k$ *if and only if* $\mathcal{G}_{\langle k\rangle} = \mathcal{X}_{\langle k\rangle} \diamondsuit_1 \mathcal{Y}_{\langle k\rangle}$.
(6) *$\mathcal{X}$ is the mode-k identity tensor if and only if $\mathcal{X}_{\langle k\rangle}$ is the mode-1 identity tensor.*
(7) *$\mathcal{X}$ is the mode-k diagonal tensor if and only if $\mathcal{X}_{\langle k\rangle}$ is the mode-1 diagonal tensor.*
(8) *$\mathcal{X}$ is the mode-k orthogonal tensor if and only if $\mathcal{X}_{\langle k\rangle}$ is the mode-1 orthogonal tensor.*

Proof. The properties **(1)**-**(4)** can be deduced straightforwardly from the provided definition. We suffice to focus on proving the last four properties. For property **(5)**, since Definition 6 (in the source document) and $\mathcal{G}_k = \mathcal{X}_k \diamondsuit_k \mathcal{Y}_k$, we have

$$\begin{aligned}
&\mathcal{G}_k = \mathcal{X}_k \diamondsuit_k \mathcal{Y}_k \\
\Leftrightarrow\ &\hat{\mathcal{G}}^{(i,j)} = \sum_{t=1}^{p} \hat{\mathcal{X}}^{(i,t)} \circledast_k \hat{\mathcal{Y}}^{(t,j)} \\
\Leftrightarrow\ &\hat{\mathcal{G}}_{mat} = \mathrm{Hbvfold}\left(\mathrm{Hbcirc}\left(\hat{\mathcal{X}}_{mat}, k\right)\mathrm{Hbvec}\left(\hat{\mathcal{Y}}_{mat}, k\right), k\right) \\
&\qquad\ \ = \mathrm{Hbvfold}\left(\mathrm{Hbcirc}\left(\hat{\mathcal{X}}_{mat,\langle k\rangle}, 1\right)\mathrm{Hbvec}\left(\hat{\mathcal{Y}}_{mat,\langle k\rangle}, 1\right), k\right),
\end{aligned}$$

where $\circledast_k$ is denoted circular convolution of $(d-2)$th-order tensors with size $n_1\times\cdots\times n_{k-1}\times n_{k+2}\times\cdots\times n_d$. The $\hat{\mathcal{G}}_{mat}$, $\hat{\mathcal{X}}_{mat}$, and $\hat{\mathcal{Y}}_{mat}$ are defined as

$$\hat{\mathcal{G}}_{mat} = \begin{bmatrix} \hat{\mathcal{G}}^{(1,1)} & \ldots & \hat{\mathcal{G}}^{(1,n_{k+1})} \\ \vdots & \ddots & \vdots \\ \hat{\mathcal{G}}^{(n_k,1)} & \ldots & \hat{\mathcal{G}}^{(n_k,n_{k+1})} \end{bmatrix}, \hat{\mathcal{X}}_{mat} = \begin{bmatrix} \hat{\mathcal{X}}^{(1,1)} & \ldots & \hat{\mathcal{X}}^{(1,n_{k+1})} \\ \vdots & \ddots & \vdots \\ \hat{\mathcal{X}}^{(n_k,1)} & \ldots & \hat{\mathcal{X}}^{(n_k,n_{k+1})} \end{bmatrix}, \hat{\mathcal{Y}}_{mat} = \begin{bmatrix} \hat{\mathcal{Y}}^{(1,1)} & \ldots & \hat{\mathcal{Y}}^{(1,n_{k+1})} \\ \vdots & \ddots & \vdots \\ \hat{\mathcal{Y}}^{(n_k,1)} & \ldots & \hat{\mathcal{Y}}^{(n_k,n_{k+1})} \end{bmatrix}. \quad (1)$$

where $\hat{\mathcal{G}}^{(i,j)} = \mathcal{G} \times_k \boldsymbol{T}_k(i,:) \times_{k+1} \boldsymbol{T}_{k+1}(j,:)$, $\hat{\mathcal{X}}^{(i,t)} = \mathcal{X} \times_k \boldsymbol{T}_k(i,:) \times_{k+1} \boldsymbol{I}_p(t,:)$, $\hat{\mathcal{Y}}^{(t,j)} = \mathcal{Y} \times_k \boldsymbol{I}_p(t,:) \times_{k+1} \boldsymbol{T}_{k+1}(j,:)$, and $\boldsymbol{I}_p \in \mathbb{R}^{p \times p}$ is the identity matrix; $\hat{\mathcal{G}}^{(i,j)}$, $\hat{\mathcal{X}}^{(i,t)}$, and $\hat{\mathcal{Y}}^{(t,j)}$ with size $n_1 \times \cdots \times n_{k-1} \times n_{k+2} \times \cdots \times n_d$ are $(d-2)$th-order tensors. Thus, $\mathrm{Hbvec}\left(\hat{\mathcal{G}}_{mat}, k\right) = \mathrm{Hbvec}\left(\hat{\mathcal{G}}_{mat,\langle k\rangle}, 1\right) = \mathrm{Hbcirc}\left(\hat{\mathcal{X}}_{mat,\langle k\rangle}, 1\right) \cdot \mathrm{Hbvec}\left(\hat{\mathcal{Y}}_{mat,\langle k\rangle}, 1\right)$. Based on property (2) of Lemma 1, we have

$$\begin{aligned}
\mathcal{G}_{mat,\langle k\rangle} &= \mathrm{Hbvfold}\left(\mathrm{Hbvec}\left(\hat{\mathcal{G}}_{mat}, k\right), 1\right) \\
&= \mathrm{Hbvfold}\left(\mathrm{Hbcirc}\left(\hat{\mathcal{X}}_{mat,\langle k\rangle}, 1\right) \cdot \mathrm{Hbvec}\left(\hat{\mathcal{Y}}_{mat,\langle k\rangle}, 1\right), 1\right) \\
\Leftrightarrow \hat{\mathcal{G}}^{(i,j)}_{mat,\langle k\rangle} &= \sum_{t=1}^{p} \hat{\mathcal{X}}^{(i,t)}_{mat,\langle k\rangle} \circledast_1 \hat{\mathcal{Y}}^{(t,j)}_{mat,\langle k\rangle} \\
\Leftrightarrow \mathcal{G}_{\langle k\rangle} &= \mathcal{X}_{\langle k\rangle} \Diamond_1 \mathcal{Y}_{\langle k\rangle}.
\end{aligned}$$

Regarding the properties **(6)** and **(7)**, they can be derived straightforwardly from Definition 9 and Definition 10 (in the source document).

As for property **(8)**, in the case where $\mathcal{X}$ is a mode-k orthogonal tensor, it satisfies

$$\mathcal{X} \Diamond_k \mathcal{X}^{\mathrm{H}_k} = \mathcal{X}^{\mathrm{H}_k} \Diamond_k \mathcal{X} = \mathcal{I}.$$

Based on the properties **(4)** and **(5)**, we can get

$$\begin{aligned}
&\mathcal{X}_{\langle k\rangle} \Diamond_1 \left(\mathcal{X}^{\mathrm{H}_k}\right)_{\langle k\rangle} = \left(\mathcal{X}^{\mathrm{H}_k}\right)_{\langle k\rangle} \Diamond_1 \mathcal{X}_{\langle k\rangle} = \mathcal{I}_{\langle k\rangle} \\
\Leftrightarrow &\mathcal{X}_{\langle k\rangle} \Diamond_1 \left(\mathcal{X}_{\langle k\rangle}\right)^{\mathrm{H}_k} = \left(\mathcal{X}_{\langle k\rangle}\right)^{\mathrm{H}_k} \Diamond_1 \mathcal{X}_{\langle k\rangle} = \mathcal{I}_{\langle k\rangle}.
\end{aligned}$$

Therefore, $\mathcal{X}_{\langle k\rangle}$ is the mode-1 orthogonal tensor. □

APPENDIX C
THE PROOF OF THEOREM 1

Theorem 1. *For $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots n_k \times p \times \cdots n_d}$ and $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots p \times n_{k+1} \times \cdots n_d}$ with the any invertible transform $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$, the mode-$k$ elt-product is equivalent to*

$$\mathcal{A} \Diamond_k \mathcal{B} = \Phi^{-1}(\Phi_{L_k}(\mathcal{A}) \triangle_k \Phi_{R_k}(\mathcal{B})). \tag{2}$$

Before the proof, we introduce some ingredients that play an important role in the proof.

Definition 3 (Full transform)**.** *For any dth-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and any invertible transform matrix $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$, where $k \in [d]$. Then the full transform of $\mathcal{X}$ is defined as*

$$\Phi(\mathcal{X}) \triangleq \mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \times \cdots \times_d \boldsymbol{T}_d^\top. \tag{3}$$

The corresponding inverse operator is defined as

$$\Phi^{-1}(\mathcal{X}) \triangleq \mathcal{X} \times_1 \left(\boldsymbol{T}_1^\top\right)^{-1} \times_2 \left(\boldsymbol{T}_2^\top\right)^{-1} \times \cdots \times_d \left(\boldsymbol{T}_d^\top\right)^{-1}. \tag{4}$$

We also note that $\Phi(\mathcal{X}) = \mathcal{X}_\Phi$ and $\Phi^{-1}(\mathcal{X}) = \mathcal{X}_{\Phi^{-1}}$.

Definition 4 (Full mode-k left and right transform)**.** *For any dth-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with any invertible transform matrix $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$, where $k \in [d]$. Then the full mode-k left transform of $\mathcal{X}$ is defined as*

$$\Phi_{L_k}(\mathcal{X}) \triangleq \mathcal{X} \times_1 \boldsymbol{T}_1^\top \cdots \times_k \boldsymbol{T}_k^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top, \tag{5}$$

and the corresponding inverse operator is defined as

$$\Phi_{L_k}^{-1}(\mathcal{X}) \triangleq \mathcal{X} \times_1 \left(\boldsymbol{T}_1^\top\right)^{-1} \cdots \times_k \left(\boldsymbol{T}_k^\top\right)^{-1} \times_{k+2} \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \cdots \times_d \left(\boldsymbol{T}_d^\top\right)^{-1}. \tag{6}$$

The full mode-k right transform of $\mathcal{X}$ is defined as

$$\Phi_{R_k}(\mathcal{X}) \triangleq \mathcal{X} \times_1 \boldsymbol{T}_1^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_{k+1} \boldsymbol{T}_{k+1}^\top \cdots \times_d \boldsymbol{T}_d^\top, \tag{7}$$

and the corresponding inverse operator is defined as

$$\Phi_{R_k}^{-1}(\mathcal{X}) \triangleq \mathcal{X} \times_1 \left(\boldsymbol{T}_1^\top\right)^{-1} \cdots \times_{k-1} \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \times_{k+1} \left(\boldsymbol{T}_{k+1}^\top\right)^{-1} \cdots \times_d \left(\boldsymbol{T}_d^\top\right)^{-1}. \tag{8}$$

We also note that $\Phi_{L_k}(\mathcal{X}) = \mathcal{X}_{\Phi_{L_k}}$ and $\Phi_{R_k}(\mathcal{X}) = \mathcal{X}_{\Phi_{R_k}}$.

Next, we give a proof of Theorem 1.

Proof. Based on the property of block circulant matrix [5, 7], for any block circulant matrix $\mathrm{Hbcirc}(\mathcal{G}, k)$, it can be block diagonalized via any invertible linear transforms $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$, i.e.,

$$\begin{aligned}\mathrm{Hbdiag}(\mathcal{G}, k) = &\left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \mathrm{Hbcirc}(\mathcal{G}, k) \\ &\left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right).\end{aligned}$$

We have

$$\begin{aligned}
&\mathcal{G} = \mathcal{A} \Diamond_k \mathcal{B} \\
&\Leftrightarrow \hat{\mathcal{G}}_{mat}^{(i,j)} = \sum_{t=1}^{p} \hat{\mathcal{A}}_{mat}^{(i,t)} \circledast_k \hat{\mathcal{B}}_{mat}^{(t,j)} \\
&\Leftrightarrow \hat{\mathcal{G}}_{mat} = \mathrm{Hbvfold}\left(\mathrm{Hbcirc}\left(\hat{\mathcal{A}}_{mat}, k\right) \mathrm{Hbvec}\left(\hat{\mathcal{B}}_{mat}, k\right), k\right) \\
&\Leftrightarrow \mathrm{Hbdiag}\left(\hat{\mathcal{G}}_{mat}, k\right) = \left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \mathrm{Hbcirc}\left(\hat{\mathcal{A}}_{mat}, k\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \mathrm{Hbcirc}\left(\hat{\mathcal{B}}_{mat}, k\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \\
&\Leftrightarrow \left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \mathrm{Hbcirc}\left(\hat{\mathcal{G}}_{mat}, k\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right) \\
&= \left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \mathrm{Hbcirc}\left(\hat{\mathcal{A}}_{mat}, k\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right) \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right) \otimes \left(\boldsymbol{T}_{k-1}^\top\right) \cdots \otimes \left(\boldsymbol{T}_1^\top\right) \otimes \boldsymbol{I}\right) \mathrm{Hbcirc}\left(\hat{\mathcal{B}}_{mat}, k\right) \\
&\left(\left(\boldsymbol{T}_d^\top\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_{k+2}^\top\right)^{-1} \otimes \left(\boldsymbol{T}_{k-1}^\top\right)^{-1} \cdots \otimes \left(\boldsymbol{T}_1^\top\right)^{-1} \otimes \boldsymbol{I}\right) \\
&\Leftrightarrow \mathrm{Hbdiag}\left(\Phi\left(\hat{\mathcal{G}}_{mat}\right) \times_k \left(\boldsymbol{T}_k^\top\right)^{-1} \times_{k+1} \left(\boldsymbol{T}_k^\top\right)^{-1}, k\right) \\
&= \left(\mathrm{Hbdiag}\left(\Phi_{L_k}\left(\hat{\mathcal{A}}_{mat}\right) \times_k \left(\boldsymbol{T}_k^\top\right)^{-1}, k\right) \cdot \mathrm{Hbdiag}\left(\Phi_{R_k}\left(\hat{\mathcal{B}}_{mat}\right) \times_{k+1} \left(\boldsymbol{T}_{k+1}^\top\right)^{-1}, k\right)\right) \\
&\Leftrightarrow \mathrm{Hbdiag}\left(\Phi\left(\mathcal{G} \times_k \left(\boldsymbol{T}_k^\top\right) \times_{k+1} \left(\boldsymbol{T}_{k+1}^\top\right)\right) \times_k \left(\boldsymbol{T}_k^\top\right)^{-1} \times_{k+1} \left(\boldsymbol{T}_k^\top\right)^{-1}, k\right) \\
&= \left(\mathrm{Hbdiag}\left(\Phi_{L_k}\left(\mathcal{A} \times_k \left(\boldsymbol{T}_k^\top\right)\right) \times_k \left(\boldsymbol{T}_k^\top\right)^{-1}, k\right) \cdot \mathrm{Hbdiag}\left(\Phi_{R_k}\left(\mathcal{B} \times_{k+1} \left(\boldsymbol{T}_{k+1}^\top\right)\right) \times_{k+1} \left(\boldsymbol{T}_{k+1}^\top\right)^{-1}, k\right)\right) \\
&\Leftrightarrow \mathrm{Hbdiag}\left(\Phi\left(\mathcal{G}\right), k\right) = \mathrm{Hbdiag}\left(\Phi_{L_k}\left(\mathcal{A}\right), k\right) \cdot \mathrm{Hbdiag}\left(\Phi_{R_k}\left(\mathcal{B}\right), k\right) \\
&\Leftrightarrow \Phi\left(\mathcal{G}\right) = \Phi_{L_k}\left(\mathcal{A}\right) \triangle_k \Phi_{R_k}\left(\mathcal{B}\right) \\
&\Leftrightarrow \mathcal{G} = \Phi^{-1}\left(\Phi_{L_k}\left(\mathcal{A}\right) \triangle_k \Phi_{R_k}\left(\mathcal{B}\right)\right),
\end{aligned}$$

where $\hat{\mathcal{G}}_{mat} = \mathcal{G} \times_k \boldsymbol{T}_k^\top \times_{k+1} \boldsymbol{T}_{k+1}^\top$, $\hat{\mathcal{A}}_{mat} = \mathcal{G} \times_k \boldsymbol{T}_k^\top \times_{k+1} \boldsymbol{I}$, and $\hat{\mathcal{B}}_{mat} = \mathcal{B} \times_k \boldsymbol{I} \times_{k+1} \boldsymbol{T}_{k+1}^\top$ based on (1). Thus, we can get

$$\mathcal{A} \Diamond_k \mathcal{B} = \mathcal{G} = \Phi^{-1}(\Phi_{L_k}(\mathcal{A}) \triangle_k \Phi_{R_k}(\mathcal{B})).$$

$\square$

APPENDIX D
THE PROOF THE THEOREM 2

Theorem 2. *For $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ with any orthogonal transform $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$ and semi-orthogonal transform $\boldsymbol{D}_k \in \mathbb{R}^{n_k \times \ell_k}$ satisfies $\boldsymbol{D}_k = \boldsymbol{T}_k(:, 1:\ell_k)$. The METNN with orthogonal transform $\boldsymbol{T}_k$ and METNN$^+$ with semi-orthogonal transform $\boldsymbol{D}_k$ have this relationship, i.e.,*

$$\|\mathcal{X}\|_{\mathrm{METNN}^+} \leq \|\mathcal{X}\|_{\mathrm{METNN}}. \tag{9}$$

Before proving this Theorem, we first give some necessary lemmas.

Lemma 2. *(Von Neuman's Trace Inequality [8]). If matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ with size $m \times n$ have singular values $\sigma_1(\boldsymbol{M}) \geq \sigma_2(\boldsymbol{M}) \geq \cdots \geq \sigma_r(\boldsymbol{M})$, and $\sigma_1(\boldsymbol{N}) \geq \sigma_2(\boldsymbol{N}) \geq \cdots \geq \sigma_r(\boldsymbol{N})$, respectively, where $r = \min(m, n)$. Then, we have*

$$\mathit{Tr}\left(\boldsymbol{M}^\top \boldsymbol{N}\right) \leq \sum_{i=1}^{r} \sigma_i\left(\boldsymbol{M}\right) \sigma_i\left(\boldsymbol{N}\right). \tag{10}$$

Specially, if $\boldsymbol{M}$ is a semi-orthogonal matrix, we can get

$$\mathit{Tr}\left(\boldsymbol{M}^\top \boldsymbol{N}\right) \leq \sum_{i=1}^{r} \sigma_i\left(\boldsymbol{M}\right) \sigma_i\left(\boldsymbol{N}\right) = \sum_{i=1}^{r} \sigma_i\left(\boldsymbol{N}\right) = \|\boldsymbol{N}\|_*. \tag{11}$$

Lemma 3. *Assuming that $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ with any orthogonal transform $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$, semi-orthogonal transforms $\boldsymbol{D}_k \in \mathbb{R}^{n_k \times \ell_k}$, and $\boldsymbol{E}_k \in \mathbb{R}^{n_k \times (n_k - \ell_k)}$ satisfy $\boldsymbol{D}_k = \boldsymbol{T}_k(:, 1:\ell_k)$ and $\boldsymbol{E}_k = \boldsymbol{T}_k(:, \ell_k + 1 : n_k)$. For any $p \neq k, k+1$, we have*

$$\left\|\|\mathrm{Hbdiag}\left(\mathcal{X} \times_p \boldsymbol{D}_p^\top\right)\|\right\|_* = \|\mathrm{Hbdiag}\left(\mathcal{X} \times_p \boldsymbol{T}_p^\top(:, \cdots, 1:\ell_p, \cdots, :)\right)\|_* \leq \|\mathrm{Hbdiag}\left(\mathcal{X} \times_p \boldsymbol{T}_p^\top\right)\|_*. \tag{12}$$

Proof. For any orthogonal transform $\boldsymbol{T}_k$ with $\boldsymbol{T}_k = [\boldsymbol{D}_k \ \boldsymbol{E}_k]$, we have

$$\begin{aligned}
&\|\mathrm{Hdiag}\left(\mathcal{X} \times_k \boldsymbol{T}_k^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_* \\
=&\|\mathrm{Hdiag}\left(\mathcal{X} \times_k [\boldsymbol{D}_k, \boldsymbol{E}_k]^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_* \\
=&\|\mathrm{Hdiag}\left(\mathcal{X} \times_k [\boldsymbol{D}_k, \mathbf{0}]^\top(:, \cdots, 1:\ell_k, \cdots, :)\right) + \mathrm{Hdiag}\left(\mathcal{X} \times_k [\mathbf{0}, \boldsymbol{E}_k]^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_*.
\end{aligned}$$

For any $\mathcal{A}$ and $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ with $\langle \mathcal{A}, \mathcal{B} \rangle = 0$, we have $\|\mathrm{Hdiag}\left(\mathcal{A} + \mathcal{B}\right)\|_* = \|\mathrm{Hdiag}\left(\mathcal{A}\right)\|_* + \|\mathrm{Hdiag}\left(\mathcal{B}\right)\|_*$. Thus, we have

$$\begin{aligned}
=&\|\mathrm{Hdiag}\left(\mathcal{X} \times_k [\boldsymbol{D}_k, \mathbf{0}]^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_* + \|\mathrm{Hdiag}\left(\mathcal{X} \times_k [\mathbf{0}, \boldsymbol{E}_k]^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_* \\
=&\|\mathrm{Hdiag}\left(\mathcal{X} \times_k \boldsymbol{D}_k^\top\right)\|_* + \|\mathrm{Hdiag}\left(\mathcal{X} \times_k [\mathbf{0}, \boldsymbol{E}_k]^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_*.
\end{aligned}$$

Since the entry of $\mathcal{X} \times_k [\mathbf{0}, \boldsymbol{E}_k]^\top(:, \cdots, i_k, \cdots, :)$ is zeros vector, the $\left\|\mathrm{Hdiag}(\mathcal{X} \times_k [\mathbf{0}, \boldsymbol{E}_k]^\top(:, \cdots, 1:\ell_k, \cdots, :))\right\|_* = \mathbf{0}$. In addition, it is obvious that

$$\|\mathrm{Hbdiag}\left(\mathcal{X} \times_k \boldsymbol{T}_k^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_* \leq \|\mathrm{Hbdiag}\left(\mathcal{X} \times_k \boldsymbol{T}_k^\top\right)\|_*.$$

Thus,

$$\|\mathrm{Hbdiag}\left(\mathcal{X} \times_k \boldsymbol{D}_k^\top\right)\|_* = \|\mathrm{Hbdiag}\left(\mathcal{X} \times_k \boldsymbol{T}_k^\top(:, \cdots, 1:\ell_k, \cdots, :)\right)\|_* \leq \|\mathrm{Hbdiag}\left(\mathcal{X} \times_k \boldsymbol{T}_k^\top\right)\|_*.$$

□

Now, we begin the proof of Theorem 2.

Proof. For $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ with any orthogonal transform $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$ and semi-orthogonal transform $\boldsymbol{D}_k \in \mathbb{R}^{n_k \times \ell_k}$ satisfies $\boldsymbol{D}_k = \boldsymbol{T}_k(:, 1:\ell_k)$. We define

$$\begin{aligned}
\|\mathcal{X}\|_{\mathrm{ETNN}_k^+} &= \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{D}_1^\top \cdots \times_k \boldsymbol{D}_k^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_*, \\
\|\mathcal{X}\|_{\mathrm{ETNN}_k} &= \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \cdots \times_k \boldsymbol{T}_k^\top \cdots \times_d \boldsymbol{T}_d^\top, k\right)\|_*.
\end{aligned}$$

For the case of $p \neq k, k+1$, based on Lemma 3, we have

$$\begin{aligned}
\|\mathcal{X}\|_{\mathrm{ETNN}_k^+} =& \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{D}_1^\top \times_2 \boldsymbol{D}_2^\top \cdots \times_{k-1} \boldsymbol{D}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{D}_{k+2}^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_* \\
\leq& \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{D}_2^\top \cdots \times_{k-1} \boldsymbol{D}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{D}_{k+2}^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_* \\
\leq& \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{D}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{D}_{k+2}^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_* \\
&\vdots \\
\leq& \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{D}_{k+2}^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_* \\
\leq& \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_* \\
&\vdots \\
\leq& \|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top, k\right)\|_*.
\end{aligned} \tag{13}$$

For the case of $p = k$ or $p = k+1$, we establish an assumption that holds for every matrix $\boldsymbol{A} \in \mathbb{R}^{n_k \times n_k}$ with transform matrix $\boldsymbol{D}_k$ and its corresponding matrix $\boldsymbol{B} = \boldsymbol{D}_k^\top \boldsymbol{A}$, where $\boldsymbol{B} \in \mathbb{R}^{\ell_k \times n}$. By employing SVD, $\boldsymbol{A}$ and $\boldsymbol{B}$ can be reformulated as follows:

$$\boldsymbol{A} = \boldsymbol{L}_{\boldsymbol{A}} \Sigma_{\boldsymbol{A}} \boldsymbol{R}_{\boldsymbol{A}}^\top, \boldsymbol{B} = \boldsymbol{L}_{\boldsymbol{B}} \Sigma_{\boldsymbol{B}} \boldsymbol{R}_{\boldsymbol{B}}^\top,$$

where $\boldsymbol{L}_{\boldsymbol{A}}, \boldsymbol{R}_{\boldsymbol{A}}, \boldsymbol{L}_{\boldsymbol{B}}$, and $\boldsymbol{R}_{\boldsymbol{B}}$ are orthogonal matrices, $\Sigma_{\boldsymbol{A}}$, and $\Sigma_{\boldsymbol{B}}$ are diagonal matrices. Thus, we can deduce that $\Sigma_{\boldsymbol{B}} = \boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{B} \boldsymbol{R}_{\boldsymbol{B}}$. Considering $\|\boldsymbol{B}\|_* = \mathrm{Tr}(\Sigma_{\boldsymbol{B}}) = \mathrm{Tr}(\boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{B} \boldsymbol{R}_{\boldsymbol{B}}) = \mathrm{Tr}(\boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{D}_k^\top \boldsymbol{A} \boldsymbol{R}_{\boldsymbol{B}})$, we can perceive that $\boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{D}_k^\top$ is a semi-orthogonal matrix, which satisfies $(\boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{D}_k^\top)(\boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{D}_k^\top)^\top = \boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{D}_k^\top \boldsymbol{D}_k \boldsymbol{L}_{\boldsymbol{B}} = \boldsymbol{I}_{\ell_k \times \ell_k}$. Base on Lemma 2, we have

$$\|\boldsymbol{B}\|_* = \mathrm{Tr}(\boldsymbol{L}_{\boldsymbol{B}}^\top \boldsymbol{D}_k^\top \boldsymbol{A} \boldsymbol{R}_{\boldsymbol{B}}) \leq \|\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{B}}\|_*.$$

Since $\boldsymbol{R}_{\boldsymbol{A}}^\top$ and $\boldsymbol{R}_{\boldsymbol{B}}$ are orthogonal, $\boldsymbol{R}_{\boldsymbol{A}}^\top \boldsymbol{R}_{\boldsymbol{B}}$ is orthogonal. The SVD of $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{B}}$ is $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{B}} = \boldsymbol{L}_{\boldsymbol{A}} \Sigma_{\boldsymbol{A}} \boldsymbol{R}_{\boldsymbol{A}}^\top \boldsymbol{R}_{\boldsymbol{B}}$. We have $\|\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{B}}\|_* = \mathrm{Tr}(\Sigma_{\boldsymbol{A}}) = \|\boldsymbol{A}\|_*$. Therefore, we have

$$\|\boldsymbol{D}_k^\top \boldsymbol{A}\|_* = \|\boldsymbol{B}\|_* \leq \|\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{B}}\|_* = \|\boldsymbol{A}\|_* = \|\boldsymbol{T}_k^\top \boldsymbol{A}\|_*. \tag{14}$$

where $\boldsymbol{T}_k$ is orthogonal. Let $\mathcal{A} = \mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top$ and $\mathcal{B} = \mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top$, we can get $\mathcal{A} = \mathcal{B} \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top$. Based on (14), we have

$$\begin{aligned}
&\|\mathrm{Hdiag}\left(\mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top\right)\|_* \\
=& \|\mathrm{Hdiag}\left(\mathcal{A}\right)\|_* = \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\mathcal{A}\left(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots\right)\|_* \\
=& \|\mathrm{Hdiag}\left(\mathcal{B} \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top\right)\|_* = \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\mathcal{B}\left(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots\right) \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top\|_* \\
=& \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\boldsymbol{D}_k^\top \boldsymbol{B}^{(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots)} \boldsymbol{D}_{k+1}\|_* \\
\leq& \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\boldsymbol{B}^{(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots)} \boldsymbol{D}_{k+1}\|_* \\
=& \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\boldsymbol{D}_{k+1}^\top \left(\boldsymbol{B}^{(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots)}\right)^\top\|_* \\
\leq& \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\boldsymbol{B}^{(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots)}\|_* \\
=& \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\boldsymbol{T}_k^\top \boldsymbol{B}^{(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots)} \boldsymbol{T}_{k+1}\|_* \\
=& \sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-1}}^{n_{k-1}} \sum_{i_{k+2}}^{n_{k+2}} \cdots \sum_{i_d}^{n_d} \|\mathcal{B}\left(i_1, \cdots, i_{k-1} :, :, i_{k+2} \cdots\right) \times_k \boldsymbol{T}_k^\top \times_{k+1} \boldsymbol{T}_{k+1}^\top\|_* = \|\mathrm{Hdiag}\left(\mathcal{B} \times_k \boldsymbol{T}_k^\top \times_{k+1} \boldsymbol{T}_{k+1}^\top\right)\|_*.
\end{aligned}$$

Therefore, by utilizing (13), we can get

$$\begin{aligned}
&\|\mathcal{X}\|_{\mathrm{ETNN}_k^+} \\
\leq&\|\mathrm{Hdiag}\left(\mathcal{X}\times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{D}_2^\top \cdots \times_{k-1} \boldsymbol{D}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{D}_{k+2}^\top \cdots \times_d \boldsymbol{D}_d^\top, k\right)\|_* \\
&\vdots \\
\leq&\|\mathrm{Hdiag}\left(\mathcal{X}\times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{D}_k^\top \times_{k+1} \boldsymbol{D}_{k+1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top\right)\|_* \\
\leq&\|\mathrm{Hdiag}\left(\mathcal{X}\times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{k-1} \boldsymbol{T}_{k-1}^\top \times_k \boldsymbol{T}_k^\top \times_{k+1} \boldsymbol{T}_{k+1}^\top \times_{k+2} \boldsymbol{T}_{k+2}^\top \cdots \times_d \boldsymbol{T}_d^\top\right)\|_* \\
=&\|\mathcal{X}\|_{\mathrm{ETNN}_k}.
\end{aligned}$$

Based on the above-mentioned analysis, we have

$$\|\mathcal{X}\|_{\mathrm{METNN}^+} = \sum_{k=1}^{d} \frac{1}{d}\|\mathcal{X}\|_{\mathrm{ETNN}_k^+} \leq \sum_{k=1}^{d} \frac{1}{d}\|\mathcal{X}\|_{\mathrm{ETNN}_k} = \|\mathcal{X}\|_{\mathrm{METNN}}.$$

□

Appendix E
The proof of Theorem 3

Theorem 3. *For any tensor $\mathcal{X}\in\mathbb{R}^{n_1\times\cdots\times n_d}$, it satisfies that $\mathcal{X}=\sum_{k=1}^{d}\frac{1}{d}\mathcal{X}_k$, $\mathcal{C}=\Phi(\mathcal{X})$, and $\mathcal{X}_k=\mathcal{U}_k\lozenge_k\mathcal{S}_k\lozenge_k\mathcal{V}_k^{\top_k}$, we have $\mathcal{C}=\sum_{k=1}^{d}\frac{1}{d}\mathcal{C}_k$ and $\mathcal{C}_k=\Phi(\mathcal{X}_k)$. Furthermore,*

$$\|\mathcal{X}\|_{\mathrm{METNN}} = \sum_{k=1}^{d}\frac{1}{d}\|\mathcal{X}\|_{\mathrm{ETNN}_k} = \sum_{k=1}^{d}\frac{1}{d}\|\mathrm{Hbdiag}\left(\mathcal{C}_{\langle k\rangle}\right)\|_*. \tag{15}$$

Proof. Based on Lemma 1, the ETNN is equivalent to

$$\begin{aligned}
\|\mathcal{X}\|_{\mathrm{ETNN}_k} &= \|\mathrm{Hbdiag}\left(\Phi\left(\mathcal{X}\right), k\right)\|_* \\
&= \|\mathrm{Hbdiag}\left(\mathcal{C}_k, k\right)\|_* \\
&= \sum_{i_1=1}^{n_1}\cdots\sum_{i_{k-1}=1}^{n_{k-1}}\sum_{i_{k+2}}^{n_{k+2}}\cdots\sum_{i_d}^{n_d}\|\mathcal{C}^{(i_1,\cdots,i_{k-1},:,:,i_{k+2},\cdots,i_d)}\|_* \\
&= \sum_{i_3}^{n_3}\sum_{i_4}^{n_4}\cdots\sum_{i_{N-1}}^{n_{N-1}}\sum_{i_d}^{n_d}\|\mathcal{C}_{\langle k\rangle}^{(:,:,i_3,i_4,\cdots,i_{N-1},i_d)}\|_* \\
&= \|\mathrm{Hbdiag}\left(\mathcal{C}_{\langle k\rangle}\right)\|_* \\
&= \|\mathcal{X}_{\langle k\rangle}\|_{\mathrm{ETNN}},
\end{aligned}$$

where we define $\|\mathcal{X}_{\langle k\rangle}\|_{\mathrm{ETNN}}\triangleq\|\mathcal{X}_{\langle k\rangle}\|_{\mathrm{ETNN},1}$. Therefore, the METNN is equivalent to

$$\begin{aligned}
\|\mathcal{X}\|_{\mathrm{METNN}} &= \sum_{k=1}^{d}\frac{1}{d}\|\mathcal{X}\|_{\mathrm{ETNN}_k} \\
&= \sum_{k=1}^{d}\frac{1}{d}\|\mathcal{X}_{\langle k\rangle}\|_{\mathrm{ETNN}} = \sum_{k=1}^{d}\frac{1}{d}\|\mathrm{Hbdiag}\left(\mathcal{C}_{\langle k\rangle}\right)\|_*.
\end{aligned}$$

where $\mathcal{X}=\sum_{k=1}^{d}\frac{1}{d}\mathcal{X}_k$. □

Appendix F
The proof of Theorem 4

Definition 5 (Tensor incoherence conditions). *For $\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ suppose that the mode-k elt-SVD is $\mathcal{X}_k=\mathcal{U}_k\lozenge_k\mathcal{S}_k\lozenge_k\mathcal{V}_k^{\top_k}$ with tensor-element multi-rank $\mathrm{rank}_{tm}(\mathcal{X})=R$. Let $\mathcal{X}=\sum_{k=1}^{d}\frac{1}{d}\mathcal{X}_k$ and $M_k=\prod_{i\neq k,k+1}^{d}n_i$. Then, for any mode-$k$, the tensor incoherence conditions with parameter $\mu>1$ are given by*

$$\max_{i_k=1,\cdots,n_l,\ l\neq k+1}\|\mathcal{U}_{\langle k\rangle}\lozenge_1\mathring{e}_k^{(i_k)}\|_F^2 \leq \frac{\mu\sum_{i=1}^{M_k}R(k,i)}{n_kM_k}, \tag{16}$$

$$\max_{i_k=1,\cdots,n_l,\ l\neq k}\|\mathcal{V}_{\langle k\rangle}\lozenge_1\mathring{e}_{k+1}^{(i_{k+1})}\|_F^2 \leq \frac{\mu\sum_{i=1}^{M_k}R(k,i)}{n_{k+1}M_k}, \tag{17}$$

where $\mathcal{U}_{\langle k\rangle}$ and $\mathcal{V}_{\langle k\rangle}$ are mode-k permutation of $\mathcal{U}_k$ and $\mathcal{V}_k$, respectively; $\mathring{e}_k^{(i_k)}$ is the standard dth-order tensor basis whose size is $n_1 \times \cdots \times n_k \times 1 \times n_{k+2} \times \cdots \times n_d$ with its $(i_1, \cdots, i_k, 1, i_{k+2}, \cdots, i_d)$-th entry be 1 and be 0 otherwise, and the $\mathring{e}_{k+1}^{(i_{k+1})} = (\mathring{e}_k^{(i_k)})^{\top_k}$.

Theorem 4. *For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with fixed orthonormal matrix $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$ and the mode-k elt-SVD of $\mathcal{X}_k = \mathcal{U}_k \lozenge_k \mathcal{S}_k \lozenge_k \mathcal{V}_k^{\mathtt{H}_k}$ with tensor-element multi-rank $\operatorname{rank}_{tm}(\mathcal{X}) = R$, they are satisfied with $\mathcal{X} = \sum_{k=1}^d \frac{1}{d} \mathcal{X}_k$, $M_k = \prod_{i \neq k, k+1}^d n_i$, $n_k^{(1)} = \max(n_k, n_{k+1})$, $n_k^{(2)} = \min(n_k, n_{k+1})$, and $\dot{n} = \max(n_1^{(1)} M_1, \cdots, n_d^{(1)} M_d)$. Suppose that the indices set $\Omega \backsim Ber(\rho)$ with $|\Omega| = m$ and the tensor incoherence conditions (16)-(17) hold. Then, there exist universal constants $c_1, c_2, c_3 > 0$ such that $\mathcal{X}$ is the unique solution with probability at least $1 - c_1 \dot{n}^{c_2}$, provided that*

$$m \geq \frac{c_3 \mu}{d} \sum_{k=1}^d \sum_{i=1}^{M_k} R(k,i)\, n_k^{(1)} \log\left(n_k^{(1)} M_k\right). \tag{18}$$

A. Main Preliminaries

In this subsection, we introduce main preliminaries and useful lemmas, which play an important role in our proof.

Definition 6. *The $\ell_{\infty,2}$-norm of the tensor $\mathcal{X}_{\langle k\rangle}$ is defined as*

$$\begin{aligned} \|\mathcal{X}_{\langle k\rangle}\|_{\infty,2} = \max\{ &\max_{(i_1,\cdots,i_k,i_{k+3},\cdots,i_d)} \|\mathcal{X}(i_1,\cdots,i_k,:,:,i_{k+3},\cdots,i_d)\|_F, \\ &\max_{(i_1,\cdots,i_{k-1},i_{k+1},i_{k+3},\cdots,i_d)} \|\mathcal{X}(i_1,\cdots,i_{k-1}:,i_{k+1},:,i_{k+3},\cdots,i_d)\|_F\}. \end{aligned} \tag{19}$$

Definition 7. *For $i_1 \in [n_1], \cdots, i_d \in [n_d]$, we define*

$$e_{k,i_1\cdots i_d} := \left(\mathring{e}_k^{(i_k)}\right) \lozenge_1 \left(\dot{e}_{k+2}^{(i_{k+2})}\right) \lozenge_1 \cdots \lozenge_1 \left(\dot{e}_d^{(i_d)}\right) \lozenge_1 \left(\dot{e}_1^{(i_1)}\right) \lozenge_1 \cdots \lozenge_1 \left(\dot{e}_{k-1}^{(i_{k-1})}\right) \lozenge_1 \left(\mathring{e}_{k+1}^{(k+1)}\right), \tag{20}$$

where an $n_1 \times \cdots \times n_d$ sized tensor with its $(i_1, \cdots, i_d)$ -th entry equaling to 1 and the rest equaling to 0; $\mathring{e}_k^{(i_k)}$ is the standard dth-order tensor basis whose size is $n_1 \times \cdots \times n_k \times 1 \times n_{k+2} \times \cdots \times n_d$ with its $(i_1, \cdots, i_k, 1, i_{k+2}, \cdots, i_d)$-th entry equaling to 1 and the rest equaling to 0, and the $\mathring{e}_{k+1}^{(i_{k+1})} = (\mathring{e}_k^{(i_k)})^{\top_k}$; $\dot{e}_{k+2}^{(i_{k+2})}$ is a tensor with the $(i_1, \cdots, i_k, 1, 1, i_{k+2}, \cdots, i_d)$-th entry of $\Phi(\dot{e}_{k+2}^{(i_{k+2})})$ equaling to $((\Phi(\vec{e}_{k+2}^{(i_{k+2})}))_{i_1,\cdots,i_k,1,1,i_{k+2},\cdots,i_d})^{-1}$ if $((\Phi(\vec{e}_{k+2}^{(i_{k+2})}))_{i_1,\cdots,i_k,1,1,i_{k+2},\cdots,i_d})^{-1} \neq 0$ and 0 otherwise.

Given any dth-order tensor $\mathcal{X}$ with the mode-k permutation of $\mathcal{X}$, i.e., $\mathcal{X}_{\langle k\rangle}$, we have

$$\begin{aligned} \mathcal{X}_{\langle k\rangle} &= \sum_{i_1\cdots i_d} \langle e_{k,i\cdots i_d}, \mathcal{X}_{\langle k\rangle}\rangle e_{k,i_1\cdots i_d}, \\ &= \sum_{i_1\cdots i_d} \mathcal{X}_{\langle k\rangle,i_1\cdots i_d} \left(\mathring{e}_k^{(i_k)}\right) \lozenge_1 \left(\dot{e}_{k+2}^{(i_{k+2})}\right) \lozenge_1 \cdots \lozenge_1 \left(\dot{e}_N^{(i_d)}\right) \lozenge_1 \left(\dot{e}_1^{(i_1)}\right) \lozenge_1 \cdots \lozenge_1 \left(\dot{e}_{k-1}^{(i_{k-1})}\right). \end{aligned} \tag{21}$$

Definition 8. *For any tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, the projection onto Ω is defined as*

$$\mathcal{P}_\Omega(\mathcal{X}) = \sum_{i_1\cdots i_d} \delta_{i_1\cdots i_d} \langle e_{i_1\cdots i_d}, \mathcal{X}\rangle e_{i_1\cdots i_d}, \tag{22}$$

where $\delta_{i_1\cdots i_d} = 1_{(i_1,\cdots,i_d)\in\Omega}$ and $1_{(\cdot)}$ denotes the indicator function. Similarly, Ω^C denotes the complementary set of Ω and $\mathcal{P}_{\Omega^\perp}$ is the projection onto Ω^C .

Definition 9. *Assume that the mode-k permutation of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is $\mathcal{X}_{\langle k\rangle}$, which its mode-$k$ elt-SVD is*

$$\mathcal{X}_{\langle k\rangle} = \mathcal{U}_{\langle k\rangle} \lozenge_1 \mathcal{S}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}. \tag{23}$$

For some $\mathcal{A}$ and $\mathcal{B}$, we define $\mathbb{T}_k$ to be the linear space

$$\mathbb{T}_k := \left\{ \mathcal{W} \,\middle|\, \mathcal{W} = \mathcal{U}_{\langle k\rangle} \lozenge_1 \mathcal{A}_{\langle k\rangle}^{\mathtt{H}_k} + \mathcal{B}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k} \right\}, \tag{24}$$

and $\mathbb{T}_k^\perp$ to be the orthogonal complement of $\mathbb{T}_k$. The orthogonal projection $\mathcal{P}_{\mathbb{T}_k}$ on $\mathbb{T}_k$ is given by

$$\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z}) = \mathcal{U}_{\langle k\rangle} \lozenge_1 \mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k} \lozenge_1 \mathcal{Z}_{\langle k\rangle} + \mathcal{Z}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k} - \mathcal{U}_{\langle k\rangle} \lozenge_1 \mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k} \lozenge_1 \mathcal{Z}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}, \tag{25}$$

and $\mathcal{P}_{\mathbb{T}_k^\perp}$ is defined as

$$\mathcal{P}_{\mathbb{T}_k^\perp}(\mathcal{Z}) = \mathcal{Z}_{\langle k\rangle} - \mathcal{P}_{\mathbb{T}_k}(\mathcal{Z}_{\langle k\rangle}) = \left(\mathcal{I} - \mathcal{U}_{\langle k\rangle} \lozenge_1 \mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k}\right) \lozenge_1 \mathcal{Z}_{\langle k\rangle} \lozenge_1 \left(\mathcal{I} - \mathcal{V}_{\langle k\rangle} \lozenge_1 \mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\right). \tag{26}$$

Lemma 4. *With the tensor $\mathcal{T} := \sum_{k=1}^d \frac{1}{d} \left(\mathcal{U}_k \lozenge_k \mathcal{V}_k^{\mathtt{H}_k}\right)$, for any mode-$k$ permutation $\mathcal{T}_{\langle k\rangle}$, we have*

$$\mathcal{T}_{\langle k\rangle} \in \mathbb{T}_k, \tag{27}$$

where the subspace $\mathbb{T}_k$ *is defined in* (24).

Lemma 5. *For any tensor* $\mathcal{X}$ *with tensor-element multi-rank* R*, and* $\mathbb{T}_k$ *be given as* (24). *Suppose that the tensor incoherence conditions* (16) *and* (17) *are satisfied, then we have*

$$\|\mathcal{P}_{\mathbb{T}_k}(e_{k,i_1\cdots i_d})\|_F^2 \leq \frac{2\mu \sum_{i=1}^{M_k} R(k,i)}{n_k^{(1)}(M_k)^2}. \tag{28}$$

Lemma 6 ([9, 10])**.** *Let* $\boldsymbol{A}$ *and* $\boldsymbol{R}$ *be given* $m \times n (m \geq n)$ *matrices. Then there exists a SVD* $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ *of* $\boldsymbol{A}$ *such that*

$$\lim_{\delta \to 0^+} \frac{\|\boldsymbol{A} + \delta \boldsymbol{R}\|_* - \|\boldsymbol{A}\|_*}{\delta} = \max_{\boldsymbol{D} \in \partial \|\boldsymbol{A}\|_*} \sum_{i=1}^{n} \boldsymbol{D}_i \boldsymbol{U}_i^\top \boldsymbol{R} \boldsymbol{V}_i, \tag{29}$$

where $\boldsymbol{D}_i, \boldsymbol{U}_i$*, and* $\boldsymbol{V}_i$ *denote the column vectors of* $\boldsymbol{D}, \boldsymbol{U}$ *and* $\boldsymbol{V}$*, respectively. Or,*

$$\lim_{\delta \to 0^+} \frac{\|\boldsymbol{A} + \delta \boldsymbol{R}\|_* - \|\boldsymbol{A}\|_*}{\delta} = \max\left\{\langle \boldsymbol{D}, \boldsymbol{R}\rangle \mid \boldsymbol{D} \in \partial \|\boldsymbol{A}\|_*, \boldsymbol{R}\right\}. \tag{30}$$

Lemma 7. *Let* $\mathcal{Q}$ *be a closed, convex subset of* $\mathbb{R}^{n_1 \times \cdots \times n_d}$*. Then* $\mathcal{Q}$ *does not contain the origin if and only if there exists* $\mathcal{G} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *such that*

$$\langle \mathcal{G}, \mathcal{K}\rangle > 0 \text{ for all } \mathcal{K} \in \mathcal{Q}. \tag{31}$$

Definition 10 ([5])**.** *Let* $\|\cdot\|_\diamond$ *be a norm on the space of* $n_1 \times \cdots \times n_d$ *real tensor. Suppose that* $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *and* $\mathcal{G} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$*, then the subdifferential (the set of subgradients) of* $\|\mathcal{X}\|_\diamond$ *is defined by*

$$\partial \|\mathcal{X}\|_\diamond = \left\{\mathcal{G} : \|\mathcal{B}\|_\diamond \geq \|\mathcal{X}\|_\diamond + \langle \mathcal{G}, \mathcal{B} - \mathcal{X}\rangle, \forall \mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}\right\}. \tag{32}$$

It can be proved that $\mathcal{G} \in \partial \|\mathcal{A}\|_\diamond$ *is equivalent to*

$$\begin{aligned} &\|\mathcal{X}\|_\diamond = \langle \mathcal{G}, \mathcal{X}\rangle, \\ &\|\mathcal{G}\|_\diamond^\ddagger \leq 1, \end{aligned} \tag{33}$$

where

$$\|\mathcal{X}\|_\diamond^\ddagger = \max_{\|\mathcal{B}\|_\diamond \leq 1} \langle \mathcal{X}, \mathcal{B}\rangle, \tag{34}$$

and $\|\cdot\|_\diamond^\ddagger$ *denotes the dual norm of* $\|\cdot\|_\diamond$.

Definition 11. *A tensor* $\mathcal{Y}_k \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *is referred to as a subgradient of* $\|\cdot\|_*$ *at* $\mathcal{M}_k$*, denoted as* $\mathcal{Y}_k \in \partial \|\mathcal{M}_k\|_*$*, if for any* $\mathcal{X}_k \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *that satisfies:*

$$\|\mathcal{X}_k\|_* - \|\mathcal{M}_k\|_* \geq \langle \mathcal{Y}_k, \mathcal{X}_k - \mathcal{M}_k\rangle. \tag{35}$$

Lemma 8. *Suppose that any tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *can be expressed as* $\mathcal{X}_{\langle k\rangle}$ *with a mode-k tensor-element rank of* R*. Its mode-k elt-SVD is given by* $\mathcal{X}_k = \mathcal{U}_{\langle k\rangle} \diamondsuit_1 \mathcal{S}_{\langle k\rangle} \diamondsuit_1 \mathcal{V}_{\langle k\rangle}^{\mathrm{H}_k}$*, then the subdifferential (the set of subgradients) of* $\|\cdot\|_*$ *at* $\mathcal{X}_k$ *can be described as:*

$$\partial \|\mathcal{X}_k\|_* = \left\{\mathcal{U}_{\langle k\rangle} \diamondsuit_1 \mathcal{V}_{\langle k\rangle}^{\mathrm{H}_k} + \mathcal{W}_k \middle| \, \mathcal{U}_{\langle k\rangle}^{\mathrm{H}_k} \diamondsuit_1 \mathcal{W} = 0, \mathcal{W}_{\langle k\rangle} \diamondsuit_1 \mathcal{V}_{\langle k\rangle} = 0, \|\mathcal{W}_k\| \leq 1\right\}. \tag{36}$$

Lemma 9. *[Matrix Bernstein Inequality [11, 12]] Consider a finite sequence* $\{\boldsymbol{Z}_k\}$ *of independent, random matrices with dimensions* $n_1 \times n_2$*. Assume that each random matrix satisfies* $\mathbb{E}\boldsymbol{Z}_k = 0$ *and* $\|\boldsymbol{Z}_k\| \leq R$ *almost surely. Define*

$$\sigma^2 := \max\left\{\left\|\sum_k \mathbb{E}\left(\boldsymbol{Z}_k \boldsymbol{Z}_k^*\right)\right\|, \left\|\sum_k \mathbb{E}\left(\boldsymbol{Z}_k^* \boldsymbol{Z}_k\right)\right\|\right\}. \tag{37}$$

Then, for all $t \geq 0$*, we have*

$$\begin{aligned} &\mathbb{P}\left\{\|\textstyle\sum_k \boldsymbol{Z}_k\| \geq t\right\} \\ &\leq (n_1 + n_2) \cdot \exp\left(\tfrac{-t^2/2}{\sigma^2 + Rt/3}\right) \\ &\leq \begin{cases} (n_1 + n_2) \cdot \exp\left(-3t^2/8\sigma^2\right) & \text{for } t \leq \sigma^2/R; \\ (n_1 + n_2) \cdot \exp(-3t/8R) & \text{for } t \geq \sigma^2/R. \end{cases} \end{aligned} \tag{38}$$

Or, for any $c > 0$*, we have*

$$\left\|\sum_k \boldsymbol{Z}_k\right\| \leq 2\sqrt{c\sigma^2 \log(n_1 + n_2)} + cR \log(n_1 + n_2), \tag{39}$$

with probability at least $1 - (n_1 + n_2)^{1-c}$.

Lemma 10. *Suppose that Ω is sampled from the Bernoulli model with parameter $\rho = \frac{m}{\prod_{i=1}^{d} n_i}$. Let $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \cdots, \times n_d}$ and Ω_k is the support Ω applied to the k-th mode. Then with the high probability*

$$\left\| \rho^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_k} \mathcal{P}_{\mathbb{T}_k} - \mathcal{P}_{\mathbb{T}_k} \right\| \leq \epsilon, \tag{40}$$

and

$$\left\| \left(\rho^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_k} \mathcal{P}_{\mathbb{T}_k} - \mathcal{P}_{\mathbb{T}_k} \right) \mathcal{Z}_{\langle k \rangle} \right\|_{\infty} \leq \epsilon \left\| \mathcal{P}_{\mathbb{T}_k} \mathcal{Z}_{\langle k \rangle} \right\|_{\infty}, \tag{41}$$

provided that $\rho \geq c_5 \epsilon^{-2} \frac{\mu \sum_{i=1}^{M_k} R(k,i) \log n_k^{(1)} M_k}{n_k^{(2)} M_k}$ for some positive numerical constant $c_5 > 0$.

Lemma 11. *Suppose that $\mathcal{Z}_{\langle k \rangle}$ is fixed, and $\Omega \sim Ber(\rho)$. Then with high probability,*

$$\left\| \left(\rho^{-1} \mathcal{P}_{\Omega^k} - \mathcal{I} \right) \mathcal{Z}_{\langle k \rangle} \right\| \leq c \left(\frac{\log(n_k^{(1)} M_k)}{\rho} \|\mathcal{Z}_{\langle k \rangle}\|_{\infty} + \sqrt{\frac{\log(n_k^{(1)} M_k)}{\rho}} \|\mathcal{Z}_{\langle k \rangle}\|_{\infty,2} \right), \tag{42}$$

provided that $\rho \geq c_4 \epsilon^{-2} \frac{\mu \sum_{i=1}^{M_k} R(k,i) \log n_k^{(1)} M_k}{n_k^{(2)} M_k}$ for some positive numerical constant $c_4 > 0$.

Lemma 12. *Suppose that $\mathcal{Z}_{\langle k \rangle}$ is fixed, and $\Omega \sim Ber(\rho)$. Then with high probability,*

$$\|\rho^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_k} \mathcal{Z}_{\langle k \rangle} - \mathcal{P}_{\mathbb{T}_k} \mathcal{Z}_{\langle k \rangle}\|_{\infty,2} \leq \frac{1}{2} \sqrt{\frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)}} \|\mathcal{Z}_{\langle k \rangle}\|_{\infty} + \frac{1}{2} \|\mathcal{Z}\|_{\infty,2}, \tag{43}$$

provided that $\rho \geq c_3 \epsilon^{-2} \frac{\mu \sum_{i=1}^{M_k} R(k,i) \log n_k^{(1)} M_k}{n_k^{(2)} M_k}$ for some positive numerical constant $c_3 > 0$.

Lemma 13. *Assume that $\left\| \rho^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}_k} - \mathcal{P}_{\mathbb{T}_k} \right\| \leq \frac{1}{2}$. Then for all $\mathcal{Z}$, we have*

$$\|\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\|_F \leq \sqrt{\frac{2}{\rho}} \left\| \mathcal{P}_{\mathbb{T}_k^{\perp}}(\mathcal{Z}) \right\|_{*}, \forall \mathcal{Z} \in \{\mathcal{Z}' : \mathcal{P}_{\Omega}(\mathcal{Z}') = 0\}. \tag{44}$$

Lemma 14. *If there exists some unfolding $k \in [K]$ such that*

$$\left\| \rho^{-1} \mathcal{P}_{T_k} \mathcal{P}_{\Gamma_k} \mathcal{P}_{T_k} - \mathcal{P}_{T_k} \right\| \leq \frac{1}{2}, \tag{45}$$

and a tensor $\mathcal{Y}$ satisfying

$$\|\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Y}_{\langle k \rangle} - \mathcal{T}_{\langle k \rangle}\right)\|_F \leq \frac{1}{4 n_k^{(1)} M_k}, \tag{46}$$

$$\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\left(\mathcal{Y}_{\langle k \rangle}\right)\|_F \leq \frac{1}{2}, \tag{47}$$

$$\mathcal{P}_{\Omega_k^{\perp}}\left(\mathcal{Y}_{\langle k \rangle}\right) = 0. \tag{48}$$

Then $\mathcal{X}_0$ is the unique solution when $n_k^{(1)} M_k$ for any $k \in [d]$ is sufficiently large.

B. Proof of Theorem 4

Proof. Consider a feasible perturbation $\mathcal{X}_0 + \triangle$, where $\mathcal{P}_{\Omega_k^{\perp}}(\triangle) = \triangle$. We now show that the objective value $f(\mathcal{X}_0 + \triangle)$ is strictly greater than $f(\mathcal{X}_0)$ unless $\triangle = 0$. Due to

$$\mathcal{U}_{0,\langle k \rangle} \lozenge_1 \mathcal{V}_{0,\langle k \rangle}^{\mathtt{H}_k} + \mathcal{W}_{0,\langle k \rangle} \in \partial \|\mathcal{X}_{0,\langle k \rangle}\|_{*}, \ for \ any \ k \in [d],$$

we can get

$$\mathcal{P}_{\mathbb{T}_k}(\mathcal{W}_{0,\langle k \rangle}) = 0, \|\mathcal{W}_{0,\langle k \rangle}\| \leq 1, \ for \ any \ k \in [d].$$

We have

$$
\begin{aligned}
&f\left(\mathcal{X}_0+\triangle\right)-f\left(\mathcal{X}_0\right)\\
&\geq \langle\sum_{k=1}^{d}\frac{1}{d}\left(\mathcal{U}_{0,k}\diamondsuit_k\mathcal{V}_{0,k}^{\mathtt{H}_k}\right)+\sum_{k=1}^{d}\frac{1}{d}\mathscr{P}^{-1}\left(\mathcal{W}_{0,\langle k\rangle}\right),\triangle\rangle\\
&=\langle\mathcal{T}_0+\sum_{k=1}^{d}\frac{1}{d}\mathscr{P}^{-1}\left(\mathcal{W}_{0,\langle k\rangle}\right),\triangle\rangle\\
&=\sum_{k=1}^{d}\frac{1}{d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*+\langle\mathcal{T}_0,\triangle\rangle\\
&=\sum_{k=1}^{d}\frac{1}{d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*+\langle\mathcal{T}_0,\triangle\rangle-\langle\mathcal{Y},\triangle\rangle &(49)\\
&=\sum_{k=1}^{d}\frac{1}{d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*+\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle-\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle\right)\\
&\geq\sum_{k=1}^{d}\frac{1}{d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*+\sum_{k=1}^{d}\frac{1}{d}\left(-\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle-\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle\right) &(50)\\
&\geq\sum_{k=1}^{d}\frac{1}{d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*-\sum_{k=1}^{d}\frac{1}{d}\left(\frac{1}{4n_k^{(1)}M_k}\|\mathcal{P}_{\mathbb{T}_k}\triangle\|_F\right)-\sum_{k=1}^{d}\frac{1}{2d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_* &(51)\\
&=\sum_{k=1}^{d}\frac{1}{2d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*-\sum_{k=1}^{d}\frac{1}{d}\left(\frac{1}{4n_k^{(1)}M_k}\|\mathcal{P}_{\mathbb{T}_k}\triangle\|_F\right)\\
&\geq\sum_{k=1}^{d}\frac{1}{d}\left(\left(\frac{1}{2}\sqrt{\frac{\rho}{2}}-\frac{1}{4n_k^{(1)}M_k}\right)\|\mathcal{P}_{\mathbb{T}_k}\triangle\|_F\right). &(52)
\end{aligned}
$$

When $n_k^{(1)}M_k$ is large enough such that

$$\frac{1}{2}\sqrt{\frac{\rho}{2}}-\frac{1}{4n_k^{(1)}M_k}>0.$$

Now, we give the proof of (49) - (51). For (49) , since $\mathcal{P}_{\Omega_k^{\perp}}\mathcal{Y}_{\langle k\rangle}=0$ and $\mathcal{P}_{\Omega_k^{\perp}}(\triangle)=\triangle$, we have

$$
\begin{aligned}
&\langle\mathcal{Y},\triangle\rangle\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{Y}_{\langle k\rangle},\triangle\rangle\right)\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\left(\mathcal{P}_{\Omega_k}+\mathcal{P}_{\Omega_k^{\perp}}\right)\left(\mathcal{Y}_{\langle k\rangle}\right),\left(\mathcal{P}_{\Omega_k}+\mathcal{P}_{\Omega_k^{\perp}}\right)(\triangle)\rangle\right)\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\Omega_k}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\Omega_k}\triangle\rangle+\langle\mathcal{P}_{\Omega_k}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\Omega_k^{\perp}}\triangle\rangle\right)\\
+&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\Omega_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\Omega_k}\triangle\rangle+\langle\mathcal{P}_{\Omega_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\Omega_k^{\perp}}\triangle\rangle\right)\\
=&0.
\end{aligned}
$$

For (50), we have

$$
\begin{aligned}
&\langle \mathcal{T}_0-\mathcal{Y},\triangle\rangle\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\left(\mathcal{P}_{\mathbb{T}_k}+\mathcal{P}_{\mathbb{T}_k^{\perp}}\right)\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\left(\mathcal{P}_{\mathbb{T}_k}+\mathcal{P}_{\mathbb{T}_k^{\perp}}\right)(\triangle)\rangle\right)\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}(\triangle)\rangle+\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle\right)\\
+&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle+\mathcal{P}_{\mathbb{T}_k^{\perp}}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\right)\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle+\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle\right)\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle+\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\mathcal{T}_{0,\langle k\rangle},\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle-\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle\right)\\
=&\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle-\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle\right).
\end{aligned}
$$

For (51), according to (46) and (47) , we have

$$
\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{0,\langle k\rangle}-\mathcal{Y}_{\langle k\rangle}\right),\mathcal{P}_{\mathbb{T}_k}\triangle\rangle\right)\geq\sum_{k=1}^{d}\frac{1}{d}\left(\frac{1}{4n_k^{(1)}M_k}\|\mathcal{P}_{\mathbb{T}_k}\triangle\|_F\right),
$$

$$
\sum_{k=1}^{d}\frac{1}{d}\left(\langle\mathcal{P}_{\mathbb{T}_k^{\perp}}\mathcal{Y}_{\langle k\rangle},\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\rangle\right)\geq\sum_{k=1}^{d}\frac{1}{2d}\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\triangle\|_*.
$$

□

The above proof assumes that the following three conditions hold, i.e.,

$$
\|\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Y}_{\langle k\rangle}-\mathcal{T}_{\langle k\rangle}\right)\|_F\leq\frac{1}{4n_k^{(1)}M_k},\tag{53}
$$

$$
\|\mathcal{P}_{\mathbb{T}_k^{\perp}}\left(\mathcal{Y}_{\langle k\rangle}\right)\|_F\leq\frac{1}{2},\tag{54}
$$

$$
\mathcal{P}_{\Omega_k^{\perp}}\left(\mathcal{Y}_{\langle k\rangle}\right)=0.\tag{55}
$$

We apply the Golfing Scheme [13] to prove these three conditions, where the architecture of proof is to refer to [3, 14].

Proof. □

We will use an approach called Golfing Scheme introduced by Gross [15] and we will follow the idea in [16, 17] where the strategy is to construct $\mathcal{Y}$ iteratively. Let Ω be a union of smaller sets Ω_k^t such that $\Omega=\bigcup_{t=1}^{t_0}\Omega_k^t$ where $t_0=\lfloor 5\log\left(n_k^{(1)}M_k\right)\rfloor$. For each t , we assume $q:=1-(1-p)^{1/t}$, and it is easy to verify that it's equivalent to our original Ω. Let $\mathcal{W}_{\langle k\rangle}^{0}=0$ and for $t=1,2,,\cdots t_0$,

$$
\mathcal{W}_{\langle k\rangle}^{t}=\mathcal{W}_{\langle k\rangle}^{t-1}+q^{-1}\mathcal{P}_{\Omega_k^t}\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{T}_{\langle k\rangle}-\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{W}_{\langle k\rangle}^{t-1}\right)\right),\tag{56}
$$

and $\mathcal{Y}_{\langle k\rangle}=\mathcal{W}_{t_0}$. By this construction we can get $\mathcal{P}_{\Omega}(\mathcal{Y}_{\langle k\rangle})=\mathcal{Y}_{\langle k\rangle}$.

For $t=0,1,\ldots,t_0$, set $\mathcal{D}_{\langle k\rangle}^{t}=\mathcal{T}_{\langle k\rangle}-\mathcal{P}_T\left(\mathcal{W}_t\right)$. Then we have $\mathcal{D}_{\langle k\rangle}^{0}=\mathcal{T}_{\langle k\rangle}$ and

$$
\mathcal{D}_{\langle k\rangle}^{t}=\left(\mathcal{P}_{\mathbb{T}_k}-q^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k^t}\mathcal{P}_{\mathbb{T}_k}\right)\left(\mathcal{D}_{\langle k\rangle}^{t-1}\right).
$$

Note that Ω_k^t is independent of $\mathcal{D}_{\langle k\rangle}^{t}$, which implies

$$
\left\|\mathcal{D}_{\langle k\rangle}^{t}\right\|_F\leq\left\|\mathcal{P}_{\mathbb{T}_k}-q^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k^t}\mathcal{P}_{\mathbb{T}_k}\right\|\left\|\mathcal{D}_{\langle k\rangle}^{t-1}\right\|_F\leq\frac{1}{2}\left\|\mathcal{D}_{\langle k\rangle}^{t-1}\right\|_F.
$$

Since $q \geq p/t_0 \geq c_5 \mu \sum_{i=1}^{M_k} R(k,i) \log\left(n_k^{(1)} M_k\right)/n$, we have

$$\begin{aligned}
&\left\|\mathcal{P}_T(\mathcal{Y}_{\langle k\rangle}) - \mathcal{T}_{\langle k\rangle}\right\|_F = \left\|\mathcal{D}_{\langle k\rangle}^{t_0}\right\|_F \\
&\leq \left(\frac{1}{2}\right)^{t_0} \left\|\mathcal{T}_{\langle k\rangle}\right\|_F \leq \frac{1}{2\left(n_k^{(1)} M_k\right)^2} \sqrt{\sum_{i=1}^{M_k} R(k,i)} \leq \frac{1}{2 n_k^{(1)} (M_k)^2},
\end{aligned}$$

holds with high probability for some large enough constants $c_5 > 0$.

From (56) we know that $\mathcal{Y}_{\langle k\rangle} = \mathcal{W}^{t_0} = \sum_{t=1}^{t_0} \left(q^{-1} \mathcal{P}_{\Omega_K^t} \mathcal{P}_T\right)\left(\mathcal{D}^{t-1}\right)$, so use Lemma 11 we obtain for some constant $c_4 > 0$,

$$\begin{aligned}
&\|\mathcal{P}_{\mathbb{T}_k^\perp}\left(\mathcal{Y}_{\langle k\rangle}\right)\|_F \\
&\leq \sum_{t=1}^{t_0} \|\mathcal{P}_{\mathbb{T}_k^\perp} q^{-1} \mathcal{P}_{\Omega_k^t} \mathcal{P}_{\mathbb{T}_k}\left(\mathcal{D}_{\langle k\rangle}^t\right)\|_F \\
&\leq \sum_{t=1}^{t_0} \|\left(q^{-1} \mathcal{P}_{\Omega_k^t} - \mathcal{I}\right) \mathcal{P}_{\mathbb{T}_k}\left(\mathcal{D}_{\langle k\rangle}^{t-1}\right)\|_F \\
&\leq c_4 \sum_{t=1}^{t_0} \left(\frac{\log\left(n_k^{(1)} M_k\right)}{m} \|\mathcal{D}_{\langle k\rangle}^{t-1}\|_\infty + \sqrt{\frac{\log\left(n_k^{(1)} M_k\right)}{m}} \|\mathcal{D}_{\langle k\rangle}^{t-1}\|_{\infty,2}\right) \\
&\leq \frac{c_4}{c_3} \sum_{t=1}^{t_0} \left(\frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)} \|\mathcal{D}_{\langle k\rangle}^{t-1}\|_\infty + \sqrt{\frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)}} \|\mathcal{D}_{\langle k\rangle}^{t-1}\|_{\infty,2}\right),
\end{aligned} \tag{57}$$

where we can bound term $\left\|\mathcal{D}_{\langle k\rangle}^{t-1}\right\|_\infty$ using Lemma 10 as follows,

$$\begin{aligned}
\|\mathcal{D}_{\langle k\rangle}^{t-1}\|_\infty &\leq \|\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_k^t} \mathcal{P}_{\mathbb{T}_k}\| \|\mathcal{D}_{\langle k\rangle}^{t-1}\|_\infty \leq \frac{1}{2} \|\mathcal{D}_{\langle k\rangle}^{t-2}\|_\infty \\
&\leq \frac{1}{2} \|\mathcal{P}_{\mathbb{T}_k} - q^{-1} \mathcal{P}_{\mathbb{T}_k} \mathcal{P}_{\Omega_k^t} \mathcal{P}_{\mathbb{T}_k}\| \|\mathcal{D}_{\langle k\rangle}^{t-3}\|_\infty \leq \left(\frac{1}{2}\right)^2 \|\mathcal{D}_{\langle k\rangle}^{t-3}\|_\infty \\
&\leq \cdots \leq \left(\frac{1}{2}\right)^{t_0-1} \|\mathcal{D}_{\langle k\rangle}^0\|_\infty = \left(\frac{1}{2}\right)^{t_0-1} \|\mathcal{T}_{\langle k\rangle}^0\|_\infty.
\end{aligned} \tag{58}$$

Based on Lemma 12, combining (57) and (58), we have

$$\begin{aligned}
&\left\|\mathcal{P}_{T^\perp}(\mathcal{Y}_{\langle k\rangle})\right\|_F \\
\leq &\frac{c_4}{\sqrt{c_3}} \frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)} \left\|\mathcal{T}_{\langle k\rangle}\right\|_\infty \sum_{t=1}^{t_0} (t+1) \left(\frac{1}{2}\right)^{t-1} + \frac{c_4}{\sqrt{c_3}} \sqrt{\frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)}} \left\|\mathcal{T}_{\langle k\rangle}\right\|_{\infty,2} \sum_{t=1}^{t_0} \left(\frac{1}{2}\right)^{t-1} \\
\leq &\frac{6c_4}{\sqrt{c_3}} \frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)} \left\|\mathcal{T}_{\langle k\rangle}\right\|_\infty + \frac{2c_4}{\sqrt{c_3}} \sqrt{\frac{n_k^{(1)} M_k}{\mu \sum_{i=1}^{M_k} R(k,i)}} \left\|\mathcal{T}_{\langle k\rangle}\right\|_{\infty,2},
\end{aligned}$$

holds with high probability for some large enough constants $c_4, c_3 > 0$. On the other hand, we have

$$\begin{aligned}
&\|\mathcal{T}_{\langle k\rangle}^0\|_{\infty,2} = \max\{\max_{(i_1,\cdots,i_k,i_{k+3},\cdots,i_d)} \|\mathcal{T}(i_1,\cdots,i_k,:,:,i_{k+3},\cdots,i_d)\|_F, \\
&\max_{(i_1,\cdots,i_{k+1},i_{k+3},\cdots,i_d)} \|\mathcal{T}(i_1,\cdots,i_{k+1},:,:,i_{k+3},\cdots,i_d)\|_F\} \\
&= \max\left\{\max_{i_1,\cdots,i_k,i_{k+3},\cdots i_d} \|\mathring{e}_{k+1}^{(k+1)} \Diamond_1 \mathcal{T}_{\langle k\rangle}^0\|_F, \max_{i_1,\cdots,i_k,i_{k+3},\cdots i_d} \|\mathring{e}_{k+2}^{(k+2)} \Diamond_1 \mathcal{T}_{\langle k\rangle}^0\|_F\right\} \\
&\leq \sqrt{\frac{\mu \sum_{i=1}^{M_k} R(k,i)}{n_k^{(1)} M_k}}.
\end{aligned}$$

It follows that with high probability

$$\left\|\mathcal{P}_{T^\perp}(\mathcal{Y}_{\langle k\rangle})\right\|_F \leq \frac{6c_4}{\sqrt{c_3}} + \frac{2c_4}{\sqrt{c_3}} \leq \frac{1}{4},$$

when c_3 large enough.

C. Proof of Lemmas

In this part, we present the proofs of some lemmas mentioned previously.

1) Proof of Lemma 4 :

Proof. For $\mathcal{X} = \mathcal{C} \times_1 \boldsymbol{T}_1 \times \cdots \times \boldsymbol{T}_N = \Phi^{-1}(\mathcal{C})$, the mode-$k$ permutation $\mathcal{X}_{\langle k \rangle} = \mathcal{C}_{\langle k \rangle} \times_1 \boldsymbol{T}_1 \times \cdots \times \boldsymbol{T}_N = \Phi^{-1}(\mathcal{X}) = \Phi^{-1}(\mathcal{C}_{\langle k \rangle})$. Assume that $\mathcal{L}_{\langle k \rangle} \Diamond_1 \Sigma_{\langle k \rangle} \Diamond_1 \mathcal{R}^{\mathtt{H}_k}_{\langle k \rangle}$ is the mode-k elt-SVD of $\mathcal{C}_{\langle k \rangle}$. Therefore, we can have

$$
\begin{aligned}
\mathcal{X}_{\langle k \rangle} &= \Phi^{-1}\left(\mathcal{C}_{\langle k \rangle}\right) = \Phi^{-1}\left(\mathcal{L}_{\langle k \rangle} \Diamond_1 \varSigma_{\langle k \rangle} \Diamond_1 \mathcal{R}^{\mathtt{H}_k}_{\langle k \rangle}\right) \\
&= \Phi^{-1}\left(\Phi^{-1}\left(\Phi\left(\mathcal{L}_{\langle k \rangle}\right) \triangle_1 \Phi\left(\varSigma_{\langle k \rangle}\right) \triangle_1 \Phi\left(\mathcal{R}^{\mathtt{H}_k}_{\langle k \rangle}\right)\right)\right) \\
&= \Phi^{-1}\Phi^{-1}\left(\Phi\Phi\Phi^{-1}\left(\mathcal{L}_{\langle k \rangle}\right) \triangle_1 \Phi\Phi\Phi^{-1}\left(\varSigma_{\langle k \rangle}\right) \triangle_1 \Phi\Phi\Phi^{-1}\left(\mathcal{R}^{\mathtt{H}_k}_{\langle k \rangle}\right)\right) \\
&= \left(\Phi^{-1}\left(\mathcal{L}_{\langle k \rangle}\right)\right) \Diamond_1 \left(\Phi^{-1}\left(\varSigma_{\langle k \rangle}\right)\right) \Diamond_1 \left(\Phi^{-1}\left(\mathcal{R}^{\mathtt{H}_k}_{\langle k \rangle}\right)\right),
\end{aligned} \tag{59}
$$

where the mode-k elt-SVD dependent transform is $\Phi\Phi(\cdot)$ in (59). Since all the transforms in Φ are orthogonal, and $\mathcal{L}_{\langle k \rangle}$ and $\mathcal{R}_{\langle k \rangle}$ are orthogonal, $\Phi^{-1}\left(\mathcal{L}_{\langle k \rangle}\right)$ and $\Phi^{-1}\left(\mathcal{R}_{\langle k \rangle}\right)$ are also orthogonal.

Therefore the subspace $\mathbb{T}_k$ in (24) corresponding to (59) is

$$
\mathbb{T}_k := \left\{ \mathcal{W} \middle| \mathcal{W} = \Phi^{-1}\left(\mathcal{L}_{\langle k \rangle}\right) \Diamond_1 \mathcal{A}^{\mathtt{H}_k}_{\langle k \rangle} + \mathcal{B}_{\langle k \rangle} \Diamond_1 \Phi^{-1}\left(\mathcal{L}_{\langle k \rangle}\right) \right\}. \tag{60}
$$

On the other hand, due to $\mathcal{X} = \Phi^{-1}(\mathcal{C})$ and $\mathcal{C} = \sum_{k=1}^{d} \frac{1}{d}\left(\mathcal{L}_k \Diamond_k \varSigma_k \Diamond_k \mathcal{R}^{\mathtt{H}_k}_k\right)$, we have

$$
\begin{aligned}
\mathcal{X} &= \Phi^{-1}\left(\mathcal{C}\right) = \Phi^{-1}\left(\sum_{k=1}^{d} \frac{1}{d}\left(\mathcal{L}_k \Diamond_k \varSigma_k \Diamond_k \mathcal{R}^{\mathtt{H}_k}_k\right)\right) \\
&= \Phi^{-1} \sum_{k=1}^{d} \frac{1}{d}\left(\Phi^{-1}\left(\Phi\left(\mathcal{L}_{\langle k \rangle}\right) \triangle_1 \Phi\left(\varSigma_{\langle k \rangle}\right) \triangle_1 \Phi\left(\mathcal{R}^{\mathtt{H}_k}_{\langle k \rangle}\right)\right)\right) \\
&= \sum_{k=1}^{d} \frac{1}{d}\left(\Phi^{-1}\Phi^{-1}\left(\Phi\Phi\Phi^{-1}\left(\mathcal{L}_k\right) \triangle_k \Phi\Phi\Phi^{-1}\left(\varSigma_k\right) \triangle_k \Phi\Phi\Phi^{-1}\left(\mathcal{R}^{\mathtt{H}_k}_k\right)\right)\right) \\
&= \sum_{k=1}^{d} \frac{1}{d}\left(\left(\Phi^{-1}\left(\mathcal{L}_k\right)\right) \Diamond_k \left(\Phi^{-1}\left(\varSigma_k\right)\right) \Diamond_k \left(\Phi^{-1}\left(\mathcal{R}^{\mathtt{H}_k}_k\right)\right)\right),
\end{aligned} \tag{61}
$$

where the transform of mode-k elt-SVD is $\Phi\Phi(\cdot)$ in (61). Therefore, the $\mathcal{T}$ can be explicitly written as

$$
\mathcal{T} = \sum_{k=1}^{d} \frac{1}{d}\left(\left(\Phi^{-1}\left(\mathcal{L}_k\right)\right) \Diamond_k \left(\Phi^{-1}\left(\mathcal{R}^{\mathtt{H}_k}_k\right)\right)\right).
$$

We define $\mathcal{H} = \sum_{k=1}^{d} \frac{1}{d}\left(\left(\Phi^{-1}\left(\mathcal{L}_k\right)\right) \Diamond_1 \left(\Phi^{-1}\left(\mathcal{R}^{\mathtt{H}_k}_k\right)\right)\right)$, and its mode-$k$ permutation is $\mathcal{H}_{\langle k \rangle}$. Let the $\mathbb{T}_k = \Phi(\mathcal{H}_{\langle k \rangle})$, we can choose $\mathcal{A}_{\langle k \rangle} = \Phi(\mathcal{L}^{\top_k}_{\langle k \rangle}) \Diamond_1 \mathcal{H}_{\langle k \rangle}$ and $\mathcal{B}_{\langle k \rangle} = 0$. Thus,

$$
\mathcal{T}_{\langle k \rangle} \in \mathbb{T}_k.
$$

□

Lemma 4 implies that the orthogonal projection of $\mathcal{Z}$ on $\mathbb{T}_k$ can be the linear space corresponding to $\mathcal{T}_{\langle k \rangle}$.

2) The proof of Lemma 5:

Proof. Due to

$$
\begin{aligned}
&\langle \mathcal{U}_{\langle k \rangle} \Diamond_1 \mathcal{U}^{\mathtt{H}_k}_{\langle k \rangle} \Diamond_1 e_{k,i_1 \cdots i_d}, e_{k,i_1 \cdots i_d} \rangle \\
=&\langle \mathcal{U}_{\langle k \rangle} \Diamond_1 \mathcal{U}^{\mathtt{H}_k}_{\langle k \rangle} \Diamond_1 \mathring{e}^{(i_k)}_k \Diamond_1 \dot{h}_k \Diamond_1 \mathring{e}^{(i_{k+1})}_{k+1}, \mathring{e}^{(i_k)}_k \Diamond_1 \dot{h}_k \Diamond_1 \mathring{e}^{(i_{k+1})}_{k+1} \rangle \\
=&\langle \mathcal{U}^{\mathtt{H}_k}_{\langle k \rangle} \Diamond_1 \mathring{e}^{(i_k)}_k, \mathcal{U}^{\mathtt{H}_k}_{\langle k \rangle} \Diamond_1 \mathring{e}^{(i_k)}_k \Diamond_1 \left(\dot{h}_k \Diamond_1 \mathring{e}^{(i_{k+1})}_{k+1} \Diamond_1 \left(\mathring{e}^{(i_{k+1})}_{k+1}\right)^{\mathtt{H}_k} \Diamond_1 \dot{h}^{\mathtt{H}_k}_k\right) \rangle \\
=&\langle \mathcal{U}^{\mathtt{H}_k}_{\langle k \rangle} \Diamond_1 \mathring{e}^{(i_k)}_k, \mathcal{U}^{\mathtt{H}_k}_{\langle k \rangle} \Diamond_1 \mathring{e}^{(i_k)}_k \rangle \\
=&\|\mathcal{U}_{\langle k \rangle} \Diamond_1 \mathring{e}^{(i_k)}_k\|_F^2,
\end{aligned}
$$

where

$$
\dot{h}_k = \left(\dot{e}^{(i_{k+2})}_{k+2}\right) \Diamond_1 \cdots \Diamond_1 \left(\dot{e}^{(i_d)}_d\right) \Diamond_1 \left(\dot{e}^{(i_1)}_1\right) \Diamond_1 \cdots \Diamond_1 \left(\dot{e}^{(i_{k-1})}_{k-1}\right).
$$

Since $\mathcal{P}_{\mathbb{T}_k}$ is self-adjoint, we have

$$\begin{aligned}
&\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\|_F^2\\
=&\langle \mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k}\Diamond_1 e_{k,i_1\cdots i_d}+e_{k,i_1\cdots i_d}\Diamond_1\mathcal{V}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k},,e_{k,i_1\cdots i_d}\rangle\\
-&\langle \mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k}\Diamond_1 e_{k,i_1\cdots i_d}\Diamond_1\mathcal{V}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k},e_{k,i_1\cdots i_d}\rangle\\
=&\|\mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k}\Diamond_1\mathring{e}_k^{(k)}\|_F^2+\|\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\Diamond_1\mathring{e}_{k+1}^{(k+1)}\|_F^2-\|\mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k}\Diamond_1\mathring{e}_k^{(k)}\Diamond_1\dot{h}_k\Diamond_1\mathring{e}_{k+1}^{(k+1)}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\|_F^2\\
\leq&\frac{\mu\sum_{i=1}^{M_k}R\left(k,i\right)\left(n_k+n_{k+1}\right)}{n_k^{(1)}n_k^{(2)}(M_k)^2}\\
\leq&\frac{2\mu\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}(M_k)^2}.
\end{aligned}$$

□

3) The proof of Lemma 7. :

Proof. Assume that $\mathbf{0}\notin\mathcal{Q}$ and consider the problem:

$$\min_{\mathcal{K}\in\mathcal{Q}}\|\mathcal{K}\|_F^2.$$

We may seek approximation from the set

$$\left\{\mathcal{K}\in\mathcal{Q}:\|\mathcal{K}\|_F^2<\left\|\mathcal{K}^{\ddagger}\right\|_F^2\right\},$$

where $\mathcal{K}^{\ddagger}\in\mathcal{Q}$ is arbitrary. Since this set is compact, the existence of a tensor $\mathcal{G}\in\mathcal{Q}$ at which the minimum is obtained is guaranteed. Now if $\mathcal{K}\in\mathcal{Q}$ is arbitrary, by the convexity of $\mathcal{Q}$, then we have

$$\lambda\mathcal{K}+(1-\lambda)\mathcal{G}\in\mathcal{Q},\quad 0\leq\lambda\leq 1.$$

Thus,

$$\begin{aligned}
0&\leq\|\lambda\mathcal{K}+(1-\lambda)\mathcal{G}\|_F^2-\|\mathcal{G}\|_F^2\\
&=\|\lambda\,\mathrm{Hbdiag}\left(\mathcal{K}_\Phi\right)+(1-\lambda)\,\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\|_F^2-\|\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\|_F^2\\
&=\left(\lambda^2\left\|\mathrm{Hbdiag}\left(\mathcal{K}_\Phi\right)-\mathrm{bdiag}\left(\mathcal{G}_\Phi\right)\right\|_F^2+2\lambda\left\langle\mathrm{Hbdiag}\left(\mathcal{K}_\Phi\right)-\mathrm{bdiag}\left(\mathcal{G}_\Phi\right),\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\right\rangle\right)\\
&=2\lambda\langle\mathcal{K}-\mathcal{G},\mathcal{G}\rangle+\lambda^2\left\|\mathrm{bdiag}\left(\mathcal{K}_\Phi\right)-\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\right\|_F^2,
\end{aligned}$$

This inequality cannot be valid for small positive λ unless

$$\langle\mathcal{K}-\mathcal{G},\mathcal{G}\rangle\geq 0,\tag{62}$$

It implies that (62) is equivalent to

$$\langle\mathcal{K},\mathcal{G}\rangle\geq\langle\mathcal{G},\mathcal{G}\rangle=\langle\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right),\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\rangle=\|\mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\|_F^2\geq 0.$$

Since $\mathcal{K}$ is arbitrary, the existence of $\mathcal{K}$ makes the conclusion hold. Now assume that there exists $\mathcal{G}\in\mathbb{R}^{n_1\times\cdots\times n_d}$ such that

$$\langle\mathcal{G},\mathcal{K}\rangle>0,\forall\mathcal{K}\in\mathcal{Q}.$$

If $\mathcal{Q}$ contains the origin, this is clearly impossible. □

4) The proof of Lemma 8. :

Proof. The Lemma 8 is proved from the following two aspects.

a) If $\mathcal{X}_k=\mathbf{0}$, then the result is obvious. Now we assume $\mathcal{X}_k\neq\mathbf{0}$ and $\mathcal{G}_k=\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}+\mathcal{W}_k$, where $\mathcal{U}_{\langle k\rangle}^{\mathtt{H}_k}\Diamond_1\mathcal{W}=\mathbf{0},\mathcal{W}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}=\mathbf{0}$ and $\|\mathcal{W}_k\|\leq 1$. It follows that

$$\begin{aligned}
\langle\mathcal{G}_k,\mathcal{X}_k\rangle&=\left\langle\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}+\mathcal{W}_k,\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{S}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\right\rangle\\
&=\left\langle\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k},\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{S}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\right\rangle+\left\langle\mathcal{W}_k,\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{S}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\right\rangle\\
&=\langle\mathrm{Hbdiag}\left(\mathcal{I}_\Phi\right),\mathrm{Hbdiag}\left(\mathcal{S}_\Phi\right)\rangle\\
&=\|\mathrm{Hbdiag}\left((\mathcal{X}_k)_\Phi\right)\|_*\\
&=\|\mathcal{X}_k\|_*.
\end{aligned}\tag{63}$$

Moreover,

$$\|\mathcal{G}_k\|=\left\|\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}+\mathcal{W}_k\right\|\leq\left\|\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k}\right\|+\|\mathcal{W}_k\|=1.\tag{64}$$

Then, equation (64) demonstrates that $\mathcal{G}_k$ is a subgradient of $\|\cdot\|_*$ at $\mathcal{X}_k$, i.e., $\mathcal{G}_k \in \partial\|\mathcal{X}_k\|_*$ (see Definition 10).

b) Let $\mathcal{F}_k := \left\{\mathcal{U}_{\langle k\rangle}\Diamond_1\mathcal{V}_{\langle k\rangle}^{\mathtt{H}_k} + \mathcal{W}_k \mid \mathcal{U}_{\langle k\rangle}^{\top_k}\Diamond_1\mathcal{W} = 0, \mathcal{W}_{\langle k\rangle}\Diamond_1\mathcal{V} = 0, \|\mathcal{W}_k\| \leq 1\right\}$. Suppose that $\mathcal{G}_k \in \partial\|\mathcal{X}_k\|_*$ but $\mathcal{G}_k \notin \mathcal{F}_k$. Then by Lemma 7 there exists $\mathcal{R} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ such that

$$\langle \mathcal{R}, \mathcal{G}_k - \mathcal{H}\rangle > 0 \text{ for all } \mathcal{H} \in \mathcal{F}_k, \mathcal{H} - \mathcal{G}_k \notin \mathbf{0},$$

so that

$$\max_{\mathcal{H}\in\mathcal{F}_k}\langle \mathcal{R}, \mathcal{H}\rangle < \max_{\mathcal{G}_k\in\partial\|\mathcal{X}_k\|_*}\langle \mathcal{R}, \mathcal{G}_k\rangle. \tag{65}$$

The equation (65) is also equivalent to

$$\max_{\mathcal{H}\in\mathcal{F}_K}\langle \mathrm{Hbdiag}\left(\mathcal{R}_\Phi\right), \mathrm{Hbdiag}\left(\mathcal{H}_\Phi\right)\rangle < \max_{\mathcal{G}\in\partial\|\mathcal{X}_k\|_*}\langle \mathrm{Hbdiag}\left(\mathcal{R}_\Phi\right), \mathrm{Hbdiag}\left((\mathcal{G}_k)_\Phi\right)\rangle.$$

Further, let $m = n_1^{(2)}M_1$. For any singular value decomposition, we have

$$\max_{\boldsymbol{D}\in\partial\|\mathrm{Hbdiag}(\mathcal{X}_{k,\Phi})\|_*}\sum_{i=1}^{m}\boldsymbol{D}_i\boldsymbol{U}_i^\top\,\mathrm{Hbdiag}\left(\mathcal{R}_\Phi\right)\boldsymbol{V}_i < \max_{\mathrm{Hbdiag}(\mathcal{G}_{k,\Phi})\in\partial\|\mathrm{Hbdiag}(\mathcal{X}_{k,\Phi})\|_*}\langle \mathrm{Hbdiag}\left(\mathcal{R}_\Phi\right), \mathrm{Hbdiag}\left(\mathcal{G}_\Phi\right)\rangle. \tag{66}$$

But the right-hand side or the left-hand side of (66) is just the standard expression for the directional derivative of the convex function $\|\mathrm{Hbdiag}\left(\mathcal{X}_{k,\Phi}\right)\|_*$ in the direction $\mathrm{Hbdiag}\left(\mathcal{R}_\Phi\right)$. Therefore, the equation (66) is contradictory to Lemma 6. The proof of $\mathcal{G}_k \in \mathcal{F}_k$ is completed.

□

5) The proof of Lemma 10. :

Proof. For any $\mathcal{Z} \in \mathbb{R}^{n_1\times n_2\cdots\times n_d}$, $\left(\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{P}_{\mathbb{T}_k} - \mathcal{P}_{\mathbb{T}_k}\right)\mathcal{Z}_{\langle k\rangle}$ can be rewrite

$$\begin{aligned}
\left(\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{P}_{\mathbb{T}_k} - \mathcal{P}_{\mathbb{T}_k}\right)\mathcal{Z}_{\langle k\rangle} &= \sum_{i_1\cdots i_d}\left(\rho^{-1}\delta_{i_1\cdots i_d} - 1\right)\langle e_{k,i_1\cdots i_d}, \mathcal{P}_{\mathbb{T}_k}\mathcal{Z}_{\langle k\rangle}\rangle\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\\
&:= \sum_{i_1\cdots i_d}\mathcal{H}_{i_1\cdots i_d}\left(\mathcal{Z}_{\langle k\rangle}\right) = \sum_{i_1\cdots i_d}\bar{\boldsymbol{H}}_{i_1\cdots i_d}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\\
&= \sum_{i_1\cdots i_d}\left(\rho^{-1}\delta_{i_1\cdots i_d} - 1\right)\langle e_{k,i_1\cdots i_d}, \mathcal{P}_{\mathbb{T}_k}\mathcal{Z}_{\langle k\rangle}\rangle\,\mathrm{Hbdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\right)\right),
\end{aligned}$$

where $\mathcal{H}_{i_1\cdots i_d} : \mathbb{R}^{n_1\times\cdots n_{k-1}\times n\times n\times n_{k+2}\times\cdots n_d} \to \mathbb{R}^{n_1\times\cdots n_{k-1}\times n\times n\times n_{k+2}\times\cdots n_d}$ is a self-adjoint random operator with $\mathbb{E}\left[\mathcal{H}_{i_1\cdots i_d}\right] = 0$. The matrix operator $\bar{\boldsymbol{H}}_{i_1\cdots i_d} : \mathbb{B} \to \mathbb{B}$, where $\mathbb{B} = \left\{\bar{\boldsymbol{B}} : \mathcal{B} \in \mathbb{R}^{n_1\times\cdots n_{k-1}\times n\times n\times n_{k+2}\times\cdots n_d}\right\}$ denotes the set consists of block diagonal matrices with the blocks as the frontal slices of $\Phi(\mathcal{B})$. Before using the Bernstein Inequality (Lemma 9) to prove the result of Lemma 10, we need to prove the boundness of $\|\mathcal{H}_{i_1\cdots i_d}\|$ and $\|\sum_{i_1\cdots i_d}\mathbb{E}\left[\mathcal{H}_{i_1\cdots i_d}^2\right]\|$. Based on the above-mentioned definitions, we have $\|\mathcal{H}_{i_1\cdots i_d}\| = \|\bar{\boldsymbol{H}}_{i_1\cdots i_d}\|$ and $\|\sum_{i_1\cdots i_d}\mathbb{E}\left[\mathcal{H}_{i_1\cdots i_d}^2\right]\| = \|\sum_{i_1\cdots i_d}\mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1\cdots i_d}^2\right]\|$. Also $\bar{\boldsymbol{H}}_{i_1\cdots i_d}$ is self-adjoint and $\mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1\cdots i_d}\right] = 0$. We only need prove the boundness of $\|\bar{\boldsymbol{H}}_{i_1\cdots i_d}\|$ and $\|\sum_{i_1\cdots i_d}\mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1\cdots i_d}^2\right]\|$.

First,

$$\begin{aligned}
\|\bar{\boldsymbol{H}}_{i_1\cdots i_d}\| = &\sup_{\|\mathrm{Hbdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\|_F=1}\|\bar{\boldsymbol{H}}_{i_1\cdots i_d}\left(\mathrm{Hdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\right)\|_F\\
\leq &\sup_{\|\mathrm{Hdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\|_F=1}\rho^{-1}\|\mathcal{Z}_{\langle k\rangle}\|_F\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\|_F\|\mathrm{Hdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\right)\right)\|_F\\
= &\sup_{\|\mathrm{Hdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\|_F=1}\rho^{-1}\|\mathrm{Hdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\|_F\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\|_F^2\\
\leq &\frac{2\mu\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}(M_k)^2},
\end{aligned}$$

where the last inequality uses (28).

Second, $\bar{\boldsymbol{H}}_{i_1\cdots i_d}^2\left(\text{Hbdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\right)=\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)^2\langle e_{k,i_1\cdots i_d},\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Z}_{\langle k\rangle}\right)\rangle\langle e_{k,i_1\cdots i_d},\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\rangle\text{Hbdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\right)\right)$. Note that $\mathbb{E}\left[\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)^2\right]\leq\rho^{-1}$, we have

$$
\begin{aligned}
&\|\sum_{i_1\cdots i_d}\mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1\cdots i_d}^2\left(\text{Hbdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\right)\right]\|_F\\
\leq&\rho^{-1}\|\sum_{i_1\cdots i_d}\langle e_{k,i_1\cdots i_d},\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Z}_{\langle k\rangle}\right)\rangle\langle e_{k,i_1\cdots i_d},\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\text{Hbdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\right)\right)\rangle\|_F\\
\leq&\rho^{-1}\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\|_F^2\|\sum_{i_1\cdots i_d}\langle e_{k,i_1\cdots i_d},\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Z}_{\langle k\rangle}\right)\rangle\|_F\\
=&\rho^{-1}\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\|_F^2\|\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Z}_{\langle k\rangle}\right)\|_F\\
\leq&\rho^{-1}\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\|_F^2\|\mathcal{Z}_{\langle k\rangle}\|_F\\
\leq&\frac{2\mu\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}M_k\rho}\|\mathcal{Z}_{\langle k\rangle}\|_F.
\end{aligned}
$$

This implies $\|\sum_{i_1\cdots i_d}\mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1\cdots i_d}^2\left(\text{Hbdiag}\left(\Phi\left(\mathcal{Z}_{\langle k\rangle}\right)\right)\right)\right]\|\leq\frac{2\mu\sum_{i=1}^{M_k}R(k,i)}{n_k^{(1)}M_k\rho}$. Let $0\leq\epsilon\leq1$, based on the Bernstein Inequality (Lemma 9), we have

$$
\begin{aligned}
&\mathbb{P}\left\{\|\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{P}_{\mathbb{T}_k}-\mathcal{P}_{\mathbb{T}_k}\|>\epsilon\right\}\\
=&\mathbb{P}\left\{\|\sum_{i_1\cdots i_d}\mathcal{H}_{i_1\cdots i_d}\|>\epsilon\right\}\\
=&\mathbb{P}\left\{\|\sum_{i_1\cdots i_d}\bar{\boldsymbol{H}}_{i_1\cdots i_d}\|>\epsilon\right\}\\
\leq&2n_k^{(1)}\rho\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2n_k^{(2)}M_k\rho}{2\mu\sum_{i=1}^{M_k}R\left(k,i\right)}\right)\\
=&2\left(n_k^{(1)}M_k\right)^{1-\frac{3}{16}c_5},
\end{aligned}
$$

where the last inequality uses $\rho\geq c_4\epsilon^{-2}\frac{\mu\sum_{i=1}^{M_k}R(k,i)\log n_k^{(1)}M_k}{n_k^{(2)}M_k}$. Thus,

$$
\left\|\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{P}_{\mathbb{T}_k}-\mathcal{P}_{\mathbb{T}_k}\right\|\leq\epsilon
$$

holds with high probability for some numerical constant c_4. □

6) The proof of Lemma 11. :

Proof. For any $\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\cdots,\times n_d}$, $\left(\rho^{-1}\mathcal{P}_{\Omega^k}-\mathcal{I}\right)\mathcal{Z}_{\langle k\rangle}$ can be rewrited

$$
\begin{aligned}
&\left(\rho^{-1}\mathcal{P}_{\Omega_k}-\mathcal{I}\right)\mathcal{Z}_{\langle k\rangle}=\sum_{i_1\cdots i_d}\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)\langle e_{k,i_1\cdots i_d},\mathcal{Z}_{\langle k\rangle}\rangle e_{k,i_1\cdots i_d}\\
=&\sum_{i_1\cdots i_d}\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}\left(\mathring{e}_k^{(i_k)}\right)\lozenge_1\left(\dot{e}_{k+2}^{(i_{k+2})}\right)\cdots\lozenge_1\left(\dot{e}_d^{(i_d)}\right)\lozenge_1\left(\dot{e}_1^{(i_1)}\right)\cdots\lozenge_1\left(\dot{e}_{k-1}^{(i_{k-1})}\right)\lozenge_1\left(\mathring{e}_{k+1}^{(k+1)}\right)\\
:=&\sum_{i_1\cdots i_d}\mathcal{G}_{\langle k\rangle,i_1\cdots i_d}.
\end{aligned}
$$

Since $\delta_{i_1\cdots i_d}$ is independent, we have $\mathbb{E}\left[\mathcal{G}_{\langle k\rangle,i_1\cdots i_d}\right]=0$ and $\|\mathcal{G}_{\langle k\rangle,i_1\cdots i_d}\|\leq\rho^{-1}\|\mathcal{Z}_{\langle k\rangle}\|$. Moreover,

$$
\begin{aligned}
&\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{G}_{\langle k\rangle,i_1\cdots i_d}\right)^\top\lozenge_1\mathcal{G}_{\langle k\rangle,i_1\cdots i_d}\right]\|\\
=&\|\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\left(\mathring{e}_{k+1}^{(i_{k+1})}\right)^\top\mathbb{E}\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)^2\|\\
=&\|\frac{1-\rho}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\left(\mathring{e}_{k+1}^{(i_{k+1})}\right)^\top\|.
\end{aligned}
$$

Based on the definition of $\mathring{e}_{k+1}^{(i_{k+1})}$, the result of $\mathring{e}_{k+1}^{(i_{k+1})} \lozenge_1 \left(\mathring{e}_{k+1}^{(i_{k+1})}\right)^\top$ is an dth-order tensor with its $(1, \cdots, 1, i_k, i_{k+1}, 1, \cdots, 1)$-th entry equaling to 1 and the rest equaling to 0. Thus, we can get

$$\begin{aligned}
&\|\mathbb{E}\left[\sum_{i_1\cdots i_d} \left(\mathcal{G}_{\langle k\rangle, i_1\cdots i_d}\right)^\top \lozenge_1 \mathcal{G}_{\langle k\rangle, i_1\cdots i_d}\right]\| \\
=&\frac{1-\rho}{\rho} \max_{i_k} \| \sum_{i_1\cdots i_{k-1} i_{k+1}\cdots i_d} |\mathcal{Z}_{\langle k\rangle, i_1\cdots i_d}|^2 \mathring{e}_k^{(i_k)} \left(\mathring{e}_k^{(i_k)}\right)^\top \| \\
\leq&\frac{1}{\rho}\|\mathcal{Z}_{\langle k\rangle}\|_{\infty,2}^2,
\end{aligned}$$

and $\|\mathbb{E}\left[\sum_{i_1\cdots i_d} \left(\mathcal{G}_{\langle k\rangle, i_1\cdots i_d}\right)^\top \lozenge_1 \mathcal{G}_{\langle k\rangle, i_1\cdots i_d}\right]\|$ is bounded. Then, using the Bernstein Inequality (Lemma 9), for any $0 \leq c \leq 1$, we have

$$\begin{aligned}
&\|\left(\rho^{-1}\mathcal{P}_{\Omega_k} - \mathcal{I}\right) \mathcal{Z}_{\langle k\rangle}\| = \| \sum_{i_1\cdots i_d} \mathcal{G}_{\langle k\rangle, i_1\cdots i_d}\| \\
\leq&\sqrt{\frac{4c'}{\rho}\|\mathcal{Z}_{\langle k\rangle}\|_{\infty,2}^2 \log\left(2n_k^{(1)} M_k\right)} + \frac{c'}{\rho}\|\mathcal{Z}_{\langle k\rangle}\|_\infty \log\left(2n_k^{(1)} M_k\right) \\
\leq& c_4 \left(\frac{\log(n_k^{(1)} M_k)}{\rho}\|\mathcal{Z}_{\langle k\rangle}\|_\infty + \sqrt{\frac{\log(n_k^{(1)} M_k)}{\rho}}\|\mathcal{Z}_{\langle k\rangle}\|_{\infty,2}\right),
\end{aligned}$$

holds with high probability for any $c_4 \geq \max\{c', 2\sqrt{c'}\}$. □

7) The proof of Lemma 12. :

Proof. For any $\mathcal{Z} \in \mathbb{R}^{n_1\times n_2\cdots,\times n_d}$, we define b-th tensor column of $\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{Z}_{\langle k\rangle} - \mathcal{P}_{\mathbb{T}_k}\mathcal{Z}_{\langle k\rangle}$ to be

$$\begin{aligned}
&\left(\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{Z}_{\langle k\rangle} - \mathcal{P}_{\mathbb{T}_k}\mathcal{Z}_{\langle k\rangle}\right) \lozenge_1 \mathring{e}^{(i_b)} \\
=&\sum_{i_1\cdots i_d} \left(\rho^{-1}\delta_{i_1\cdots i_d} - 1\right) \mathcal{Z}_{\langle k\rangle, i_1\cdots i_d} \mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right) \lozenge_1 \mathring{e}_k^{(i_b)} \\
:=&\sum_{i_1\cdots i_d} \mathcal{K}_{k,i_1\cdots i_d},
\end{aligned}$$

where $\mathcal{K}_{k,i_1\cdots i_d} \in \mathbb{R}^{n_1\times\cdots\times n_k\times 1\times n_{k+2}\times\cdots\times n_d}$ are zero-mean independent tensor columns. Based on Lemma 5, we have

$$\|\mathcal{K}_{k,i_1\cdots i_d}\|_F = \|\left(\rho^{-1}\delta_{i_1\cdots i_d} - 1\right) \mathcal{Z}_{\langle k\rangle, i_1\cdots i_d} \mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right) \lozenge_1 \mathring{e}_k^{(i_b)}\|_F \leq \rho^{-1}\sqrt{\frac{2\mu\sum_{i=1}^{M_k} R(k,i)}{n_k^{(1)} M_k}}\|\mathcal{Z}_{\langle k\rangle}\|_\infty.$$

Moreover,

$$
\begin{aligned}
&\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{K}_{k,i_1\cdots i_d}\right)^{\top}\lozenge_1\mathcal{K}_{k,i_1\cdots i_d}\right]\|_F\\
=&\|\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\left(\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\lozenge_1\mathring{e}_k^{(i_b)}\right)^{\top}\lozenge_1\left(\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\lozenge_1\mathring{e}_k^{(i_b)}\right)\mathbb{E}\left(\rho^{-1}\delta_{i_1\cdots i_d}\right)^2\|_F\\
=&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\|\mathcal{P}_{\mathbb{T}_k}\left(e_{k,i_1\cdots i_d}\right)\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2\\
=&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\|\mathcal{U}_{\langle k\rangle}\lozenge_1\mathcal{U}_{\langle k\rangle}^{\top}\lozenge_1 e_{k,i_1\cdots i_d}\lozenge_1\mathring{e}_k^{(i_b)}+\left(\mathcal{I}-\mathcal{U}_{\langle k\rangle}\lozenge_1\mathcal{U}_{\langle k\rangle}^{\top}\right)\lozenge_1 e_{k,i_1\cdots i_d}\lozenge_1\mathcal{V}_{\langle k\rangle}\lozenge_1\mathcal{V}_{\langle k\rangle}^{\top}\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2\\
\leq&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\left(\frac{\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}M_k\mu^{-1}}\|\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2+\|\left(\mathcal{I}-\mathcal{U}_{\langle k\rangle}\lozenge_1\mathcal{U}_{\langle k\rangle}^{\top}\lozenge_1\mathring{e}_k^{(i_k)}\lozenge_1\mathring{e}_k^{(i_b)}\right)\|_F^2\|\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\mathcal{V}_{\langle k\rangle}\lozenge_1\mathcal{V}_{\langle k\rangle}^{\top}\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2\right)\\
\leq&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\left(\frac{\mu\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}M_k}\|\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2+\|\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\mathcal{V}_{\langle k\rangle}\lozenge_1\mathcal{V}_{\langle k\rangle}^{\top}\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2\right)\\
\leq&\frac{\mu\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}M_k\rho}\|\mathcal{Z}\|_{\infty,2}^2+\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{\langle k\rangle,i_1\cdots i_d}|^2\|\mathring{e}_{k+1}^{(i_{k+1})}\lozenge_1\mathcal{V}_{\langle k\rangle}\lozenge_1\mathcal{V}_{\langle k\rangle}^{\top}\lozenge_1\mathring{e}_k^{(i_b)}\|_F^2\\
\leq&\frac{2\mu\sum_{i=1}^{M_k}R\left(k,i\right)}{n_k^{(1)}M_k\rho}\|\mathcal{Z}\|_{\infty,2}^2.
\end{aligned}
$$

So, $\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{K}_{k,i_1\cdots i_d}\right)^{\top}\lozenge_1\mathcal{K}_{k,i_1\cdots i_d}\right]\|_F$ is bounded. Based on Bernstein Inequality (Lemma 9), we have

$$\|\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega_k}\mathcal{Z}_{\langle k\rangle}-\mathcal{P}_{\mathbb{T}_k}\mathcal{Z}_{\langle k\rangle}\|_{\infty,2}\leq\frac{1}{2}\sqrt{\frac{n_k^{(1)}M_k}{\mu\sum_{i=1}^{M_k}R\left(k,i\right)}}\|\mathcal{Z}_{\langle k\rangle}\|_{\infty}+\frac{1}{2}\|\mathcal{Z}\|_{\infty,2},$$

holds with high probability for $m\geq c_3\mu\epsilon^{-2}\sum_{i=1}^{M_k}R\left(k,i\right)n_k^{(1)}\log\left(n_k^{(1)}M_k\right)$. □

8) *The proof of Lemma 13. :*

Proof. To prove this Lemma, we can start from the following two aspects. On the one hand, we deduce

$$
\begin{aligned}
&\left\|\rho^{-1}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\right\|_F^2\\
&=\left\langle\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z}),\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\right\rangle\\
&=\left\langle\left(\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k}-\mathcal{P}_{\mathbb{T}_k}\right)(\mathcal{Z}),\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\right\rangle+\langle\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z}),\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\rangle\\
&\geq\|\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\|_F^2-\left\|\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k}-\mathcal{P}_{\mathbb{T}_k}\right\|\|\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\|_F^2\\
&\geq\tfrac{1}{2}\|\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\|_F^2,
\end{aligned}
$$

where the last inequality utilizes the given condition: $\left\|\rho^{-1}\mathcal{P}_{\mathbb{T}_k}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k}-\mathcal{P}_{\mathbb{T}_k}\right\|\leq\frac{1}{2}$. On the other hand, the condition $\mathcal{P}_{\Omega}(\mathcal{Z})=0$ deduces that $\mathcal{P}_{\Omega}(\mathcal{Z})=0$, and thus

$$\frac{1}{\sqrt{2}}\|\mathcal{P}_{\mathbb{T}_k}\left(\mathcal{Z}\right)\|_F\leq\frac{1}{\sqrt{\rho}}\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\|_F=\frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}_k^{\perp}}(\mathcal{Z})\right\|_F\leq\frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\mathbb{T}_k^{\perp}}(\mathcal{Z})\right\|_F\leq\frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\mathbb{T}_k^{\perp}}(\mathcal{Z})\right\|_*,$$

where the last inequality utilizes

$$\|\mathcal{Z}\|_F=\|\mathrm{Hbdiag}\left(\mathcal{Z}\right)\|_F\leq\|\mathrm{Hbdiag}\left(\mathcal{Z}\right)\|_*=\|\mathcal{Z}\|_*.$$

Thus,

$$\|\mathcal{P}_{\mathbb{T}_k}(\mathcal{Z})\|_F\leq\sqrt{\frac{2}{\rho}}\left\|\mathcal{P}_{\mathbb{T}_k^{\perp}}(\mathcal{Z})\right\|_*.$$

□

Appendix G
Convergence Analysis

In this section, we provide the theoretical convergence of our algorithm. The architecture of proof is to refer to [18], where the proposed algorithm satisfies the following conditions:

1) The sequence $(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t)_{t\in\mathbb{N}}$ satisfies sufficient decrease condition;
2) The sequence $(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t)_{t\in\mathbb{N}}$ is bounded;
3) The $\mathcal{L}(\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k, \Lambda)$ is a proper lower semi-continuous function and has the Kurdyka-Łojasiewicz (K-Ł) property [19] at $(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t)_{t\in\mathbb{N}}$;
4) The $\mathcal{L}(\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k, \Lambda)$ satisfies subgradient bound condition.

Theorem 5. *Suppose the sequence $(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t)_{t\in\mathbb{N}}$ is generated by the proposed algorithm, it satisfies the conditions 1-4. Then, the sequence converges to a critical point of $\mathcal{L}(\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k, \Lambda)$.*

Before proving Theorem 5, we first define auxiliary functions:

$$\begin{aligned}\mathcal{L}_\beta\left(\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k, \Lambda\right) = \sum_{k=1}^{d} \frac{1}{d}\left\{\|\mathrm{Hbdiag}\left(\mathcal{Z}_k, k\right)\|_* + \langle \Gamma_k, \mathcal{C}_k - \mathcal{Z}_k\rangle + \frac{\beta}{2}\|\mathcal{C}_k - \mathcal{Z}_k\|_F^2\right\} + \Phi\left(\mathcal{X}\right) + \Psi\left(\boldsymbol{T}_k\right) \\ + \langle \Lambda, \mathcal{X} - \mathcal{C} \times_1 \boldsymbol{T}_1 \times_2 \cdots \times_d \boldsymbol{T}_d\rangle + \frac{\beta}{2}\|\mathcal{X} - \mathcal{C} \times_1 \boldsymbol{T}_1 \times_2 \cdots \times_d \boldsymbol{T}_d\|_F^2,\end{aligned} \tag{67}$$

$$\mathcal{L}_1\left(\mathcal{C}_k, \mathcal{Z}_k, \Gamma_k\right) = \|\mathrm{Hbdiag}\left(\mathcal{Z}_k, k\right)\|_* + \langle \Gamma_k, \mathcal{C}_k - \mathcal{Z}_k\rangle + \frac{\beta}{2}\|\mathcal{C}_k - \mathcal{Z}_k\|_F^2, \tag{68}$$

$$\mathcal{L}_2\left(\mathcal{X}, \mathcal{C}, \boldsymbol{T}_k, \Lambda\right) = \langle \Lambda, \mathcal{X} - \mathcal{C} \times_1 \boldsymbol{T}_1 \times_2 \cdots \times_d \boldsymbol{T}_d\rangle + \frac{\beta}{2}\|\mathcal{X} - \mathcal{C} \times_1 \boldsymbol{T}_1 \times_2 \cdots \times_d \boldsymbol{T}_d\|_F^2, \tag{69}$$

$$\mathcal{L}_3\left(\mathcal{Z}_k\right) = \|\mathrm{Hbdiag}\left(\mathcal{Z}_k, k\right)\|_*. \tag{70}$$

where $\mathcal{L}_\beta\left(\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k, \Lambda\right)$ is the objective function. Within the framework of ADMM, $\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k$, and Λ are alternately updated as

$$\left\{\begin{array}{l} \mathcal{X}^{t+1} = \arg\min\limits_{\mathcal{X}} \langle \Lambda^t, \mathcal{X} - \mathcal{C}^t \times_1 \boldsymbol{T}_1^t \times_2 \cdots \times_d \boldsymbol{T}_d^t\rangle + \frac{\beta}{2}\|\mathcal{X} - \mathcal{C}^t \times_1 \boldsymbol{T}_1^t \times_2 \cdots \times_d \boldsymbol{T}_d^t\|_F^2 + \Phi\left(\mathcal{X}\right), \\ \mathcal{C}^{t+1} = \arg\min\limits_{\mathcal{C}}\ \sum_{k=1}^{d} \frac{1}{d}\left\{\langle \Gamma_k^t, \mathcal{C}_k - \mathcal{Z}_k^t\rangle + \frac{\beta}{2}\|\mathcal{C}_k - \mathcal{Z}_k^t\|_F^2\right\} + \frac{\beta}{2}\|\mathcal{X}^{t+1} - \mathcal{C} \times_1 \boldsymbol{T}_1^t \times_2 \cdots \times_d \boldsymbol{T}_d^t + \frac{\Lambda^t}{\beta}\|_F^2, \\ \mathcal{Z}_k^{t+1} = \arg\min\limits_{\mathcal{Z}_k}\ \|Hb\mathrm{diag}\left(\mathcal{Z}_k, k\right)\|_* + \langle \Gamma_k^t, \mathcal{C}_k^{t+1} - \mathcal{Z}_k\rangle + \frac{\beta}{2}\|\mathcal{C}_k^{t+1} - \mathcal{Z}_k\|_F^2, \\ \boldsymbol{T}_k^{t+1} = \arg\min\limits_{\boldsymbol{T}_k}\ \langle \Lambda^t, \mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1 \times_2 \cdots \times_d \boldsymbol{T}_d\rangle + \frac{\beta}{2}\|\mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1 \times_2 \cdots \times_d \boldsymbol{T}_d\|_F^2 + \Psi\left(\boldsymbol{T}_k\right), \\ \Gamma_k^{t+1} = \Gamma_k^t + \beta\left(\mathcal{C}_k^{t+1} - \mathcal{Z}_k^{t+1}\right), \\ \Lambda^{t+1} = \Lambda^t + \beta\left(\mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \cdots \times_d \boldsymbol{T}_d^{t+1}\right), \end{array}\right. \tag{71}$$

where $k = 1, 2, \cdots, d$.

Lemma 15 (Sufficient decrease lemma)**.** *The sequence $(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t)_{t\in\mathbb{N}}$ be generated by* (71) *satisfies the following formulae:*

$$\left\{\begin{array}{l} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}, \Gamma^{t+1}\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}, \Gamma^t\right). \end{array}\right. \tag{72}$$

Proof. Let $\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}$, and Γ^{t+1} are minimizers of $\mathcal{L}(\mathcal{X}, \mathcal{C}, \mathcal{Z}_k, \boldsymbol{T}_k, \Gamma_k, \Lambda)$. Then, we have

$$\left\{\begin{array}{l} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^t, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^t, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}, \Gamma^t\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^t, \Gamma^t\right), \\ \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}, \Gamma^{t+1}\right) \le \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Lambda_k^{t+1}, \Gamma^t\right). \end{array}\right. \tag{73}$$

□

Lemma 16 (Boundedness)**.** *The sequence $(\mathcal{X}^t, \mathcal{C}^t, \mathcal{Z}_k^t, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t)_{t\in\mathbb{N}}$ be generated by* (71) *is bounded.*

Proof. According to Lemma 15, we can conclude that

$$\begin{aligned}\|\mathrm{Hbdiag}\left(\mathcal{Z}_k^t,k\right)\|_* = \mathcal{L}_2\left(\mathcal{Z}_k^t\right) &\leq \mathcal{L}\left(\mathcal{X}^{t+1},\mathcal{C}^{t+1},\mathcal{Z}_k^{t+1},\boldsymbol{T}_k^t,\Lambda_k^t,\Gamma^t\right) \leq \mathcal{L}\left(\mathcal{X}^{t+1},\mathcal{C}^{t+1},\mathcal{Z}_k^{t},\boldsymbol{T}_k^t,\Lambda_k^t,\Gamma^t\right)\\ &\leq \cdots \leq \mathcal{L}\left(\mathcal{X}^{0},\mathcal{C}^{0},\mathcal{Z}_k^{0},\boldsymbol{T}_k^0,\Lambda_k^0,\Gamma^0\right).\end{aligned} \tag{74}$$

Since $\mathcal{L}\left(\mathcal{X}^{0},\mathcal{C}^{0},\mathcal{Z}_k^{0},\boldsymbol{T}_k^0,\Lambda_k^0,\Gamma^0\right)$ is a constant, $\|\mathrm{Hbdiag}\left(\mathcal{Z}_k^t,k\right)\|_*$ is bound. From the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|\mathcal{Z}_k^t\|_F^2 &\leq \sum_{i_1=1}^{n_1}\cdots\sum_{i_{k-1}}^{n_{k-1}}\sum_{i_{k+2}}^{n_{k+2}}\cdots\sum_{i_d}^{n_d}\|\mathcal{Z}\left(i_1,\cdots,i_{k-1},:,:,i_{k+2},\cdots,i_d\right)^t\|_F^2\\ &\leq \sum_{i_1=1}^{n_1}\cdots\sum_{i_{k-1}}^{n_{k-1}}\sum_{i_{k+2}}^{n_{k+2}}\cdots\sum_{i_d}^{n_d}\|\mathcal{Z}\left(i_1,\cdots,i_{k-1},:,:,i_{k+2},\cdots,i_d\right)^t\|_* = \|\mathrm{Hbdiag}\left(\mathcal{Z}_k^t,k\right)\|_*.\end{aligned} \tag{75}$$

Thus, $\mathcal{Z}_k^t$ is bound. Since $\boldsymbol{T}_k^t = \boldsymbol{U}_k^t\boldsymbol{V}_k^{\top t}$, we can have

$$\|\boldsymbol{T}_k^t\|_F^2 \leq \|\boldsymbol{U}_k^t\|_F^2\|\boldsymbol{V}_k^t\|_F^2 \leq \sqrt{r_i}\sqrt{n_i}, \tag{76}$$

where $\boldsymbol{U}_k^t$ and $\boldsymbol{V}_k^t$ are semi-orthogonal. Thus, $\boldsymbol{T}_k^t$ is bound. Due to the triangle inequality, we have

$$\|\mathcal{C}_k^t\|_F^2 - \|\mathcal{Z}_k^t\|_F^2 \leq \|\mathcal{C}_k^t - \mathcal{Z}_k^t\|_F^2, \tag{77}$$

which is equivalent to

$$\|\mathcal{C}_k^t\|_F^2 \leq \|\mathcal{Z}_k^t\|_F^2 + \|\mathcal{C}_k^t - \mathcal{Z}_k^t\|_F^2. \tag{78}$$

Suppose $\|\mathcal{C}_k^t - \mathcal{Z}_k^t\|_F^2 \to +\infty$, we have

$$\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k^t,\Lambda^t) \to +\infty. \tag{79}$$

This result contradicts Lemma 15, which emphasizes $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k^t,\Lambda^t)$ is bound. Thus, $\|\mathcal{C}_k^t - \mathcal{Z}_k^t\|_F^2$ is bound. And because $\|\mathcal{Z}_k^t\|_F^2$ is bounded, we can get $\|\mathcal{C}_k^t\|_F^2$ is bound via (78). Similarly, we have

$$\begin{aligned}\|\mathcal{X}^t\|_F^2 - \|\mathcal{C}_k^t\|_F^2\|\boldsymbol{T}_1^t\|_F^2\cdots\|\boldsymbol{T}_d^t\|_F^2 &\leq \|\mathcal{X}^t\| - \|\mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\|_F^2\\ &\leq \|\mathcal{X}^t - \mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\|_F^2,\end{aligned} \tag{80}$$

which is equivalent to

$$\|\mathcal{X}^t\|_F^2 \leq \|\mathcal{C}_k^t\|_F^2\|\boldsymbol{T}_1^t\|_F^2\cdots\|\boldsymbol{T}_d^t\|_F^2 + \|\mathcal{X}^t - \mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\|_F^2. \tag{81}$$

Suppose $\|\mathcal{X}^t - \mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\|_F^2 \to +\infty$, we have

$$\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k^t,\Lambda^t) \to +\infty. \tag{82}$$

This result contradicts Lemma 15, which emphasizes $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k^t,\Lambda^t)$ is bound. Thus, $\|\mathcal{X}^t - \mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\|_F^2$ is bound. So, $\|\mathcal{X}^t\|_F^2$ is bound. In the optimization problem (71), the $\mathcal{C}_k^{t+1}$ and $\mathcal{Z}_k^{t+1}$ are optimal solutions and they satisfy

$$\begin{cases}\mathcal{L}_1\left(\mathcal{C}_k^{t+1},\mathcal{Z}_k^{t+1},\Gamma_k^{t+1},\beta^{t+1}\right) \leq \mathcal{L}_1\left(\mathcal{C}_k^{t+1},\mathcal{Z}_k^{t+1},\Gamma_k^{t},\beta^{t}\right)\\ \Gamma_k^{t+1} = \Gamma_k^t + \beta^t\left(\mathcal{C}_k^{t+1} - \mathcal{Z}_k^{t+1}\right).\end{cases} \tag{83}$$

Then, we have

$$\begin{aligned}&\mathcal{L}_1\left(\mathcal{C}_k^{t+1},\mathcal{Z}_k^{t+1},\Gamma_k^{t+1},\beta^{t+1}\right) - \mathcal{L}_1\left(\mathcal{C}_k^{t+1},\mathcal{Z}_k^{t+1},\Gamma_k^{t},\beta^{t}\right)\\ &= \langle\Gamma_k^{t+1} - \Gamma_k^t, \mathcal{C}_k^{t+1} - \mathcal{Z}_k^{t+1}\rangle + \frac{\beta^{t+1}-\beta^t}{2}\|\mathcal{C}_k^{t+1} - \mathcal{Z}_k^{t+1}\|_F^2\\ &= \langle\Gamma_k^{t+1} - \Gamma_k^t, \frac{\Gamma_k^{t+1} - \Gamma_k^t}{\beta^t}\rangle + \frac{\beta^{t+1}-\beta^t}{2}\|\frac{\Gamma_k^{t+1} - \Gamma_k^t}{\beta^t}\|_F^2\\ &= \frac{\beta^{t+1}+\beta^t}{2\left(\beta^t\right)^2}\|\Gamma_k^{t+1} - \Gamma_k^t\|_F^2\end{aligned} \tag{84}$$

Based on (84), we easily obtain

$$\mathcal{L}_1\left(\mathcal{C}_k^{t+1},\mathcal{Z}_k^{t+1},\Gamma_k^{t+1},\beta^{t+1}\right) \leq \mathcal{L}_1\left(\mathcal{C}_k^{0},\mathcal{Z}_k^{0},\Gamma_k^{0},\beta^{0}\right) + \sum_{t=1}^{\infty}\frac{\beta^{t+1}+\beta^t}{2\left(\beta^t\right)^2}\|\Gamma_k^{t+1} - \Gamma_k^t\|_F^2 \tag{85}$$

For the penalty parameter β, it satisfies

$$\sum_{k=1}^{\infty}\frac{\beta^{t+1}+\beta^t}{2\left(\beta^t\right)^2} \leq \sum_{k=1}^{\infty}\frac{\beta^{t+1}}{\left(\beta^t\right)^2} \leq +\infty, \tag{86}$$

which implies that $\sum_{k=1}^{\infty} \frac{\beta^{t+1}+\beta^{t}}{2(\beta^{t})^{2}}$ is convergent. Due to the $\mathcal{L}_1\left(\mathcal{C}_k, \mathcal{Z}_k, \mathit{\Gamma}_k, \beta\right)$ is bound, the $\sum_{t=1}^{\infty} \frac{\beta^{t+1}+\beta^{t}}{2(\beta^{t})^{2}} \|\mathit{\Gamma}_k^{t+1} - \mathit{\Gamma}_k^{t}\|_F^2$ is bound. And because the $\sum_{k=1}^{\infty} \frac{\beta^{t+1}+\beta^{t}}{2(\beta^{t})^{2}}$ is convergent, the $\|\mathit{\Gamma}_k^{t+1} - \mathit{\Gamma}_k^{t}\|_F^2$ is monotone and bounded. Thus, $\|\mathit{\Gamma}_k^{t}\|_F^2$ is bound. Similarly, In the optimization problem (71), the $\mathcal{X}^{t+1}, \mathcal{C}^{t+1}$ and $\boldsymbol{T}_k^{t+1}$ are optimal solutions and they satisfy

$$\left\{ \begin{array}{l} \mathcal{L}_2\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Lambda}^{t+1}, \beta^{t+1}\right) \leq \mathcal{L}_2\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Lambda}^{t}, \beta^{t}\right) \\ \mathit{\Lambda}^{t+1} = \mathit{\Lambda}^{t} + \beta\left(\mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \cdots \times_d \boldsymbol{T}_d^{t+1}\right). \end{array} \right. \tag{87}$$

Then, we have

$$\begin{aligned} & \mathcal{L}_2\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Lambda}^{t+1}, \beta^{t+1}\right) - \mathcal{L}_2\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Lambda}^{t}, \beta^{t}\right) \\ & = \langle \mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t}, \mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \times_2 \cdots \times_d \boldsymbol{T}_d^{t+1} \rangle + \frac{\beta^{t+1} - \beta^{t}}{2} \|\mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \times_2 \cdots \times_d \boldsymbol{T}_d^{t+1}\|_F^2 \\ & = \langle \mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t}, \frac{\mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t}}{\beta^{t}} \rangle + \frac{\beta^{t+1} - \beta^{t}}{2} \|\frac{\mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t}}{\beta^{t}}\|_F^2 \\ & = \frac{\beta^{t+1} + \beta^{t}}{2\left(\beta^{t}\right)^2} \|\mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t}\|_F^2 \end{aligned} \tag{88}$$

For the penalty parameter β, it satisfies

$$\sum_{k=1}^{\infty} \frac{\beta^{t+1} + \beta^{t}}{2\left(\beta^{t}\right)^2} \leq \sum_{k=1}^{\infty} \frac{\beta^{t+1}}{\left(\beta^{t}\right)^2} \leq +\infty. \tag{89}$$

Similar to the proof process of $\|\mathit{\Gamma}_k^{t}\|_F^2$, $\|\mathit{\Lambda}^{t}\|_F^2$ is bound.

In summary, the sequence $(\mathcal{X}^{t}, \mathcal{C}^{t}, \mathcal{Z}_k^{t}, \boldsymbol{T}_k^{t}, \mathit{\Gamma}_k^{t}, \mathit{\Lambda}^{t})_{t\in\mathbb{N}}$ is bounded. Therefore, the proof of Lemma 16 is completed.

□

Lemma 17 (Subgradient bound)**.** *The sequence* $(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Gamma}_k^{t+1}, \mathit{\Lambda}^{t+1})_{t\in\mathbb{N}}$ *is generated by* (71)*, and there exist* $C(\beta) > 0$ *and* $\boldsymbol{D}_w^{t+1}$ $(w = 1, \cdots, 4)$*. For any* $\boldsymbol{D}_w^{t+1}$*, it satisfies the following formulae*

$$\|\boldsymbol{D}_w^{t+1}\|_F^2 \leq 2C\left(\beta\right)\left(\|\mathcal{X}^{t+1} - \mathcal{X}^{t}\|_F^2 + \|\mathcal{C}^{t+1} - \mathcal{C}^{t}\|_F^2 + \sum_{k=1}^{d} \frac{1}{d}\left(\|\mathcal{Z}_k^{t+1} - \mathcal{Z}_k^{t}\|_F^2 + \|\boldsymbol{T}_k^{t+1} - \boldsymbol{T}_k^{t}\|_F^2\right)\right), \tag{90}$$

where

$$\left\{ \begin{array}{l} \boldsymbol{D}_1^{t+1} \in \partial_{\mathcal{X}} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Gamma}_k^{t+1}, \mathit{\Lambda}^{t+1}\right), \\ \boldsymbol{D}_2^{t+1} \in \partial_{\mathcal{C}} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Gamma}_k^{t+1}, \mathit{\Lambda}^{t+1}\right), \\ \boldsymbol{D}_3^{t+1} \in \partial_{\mathcal{Z}_k} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Gamma}_k^{t+1}, \mathit{\Lambda}^{t+1}\right), \\ \boldsymbol{D}_4^{t+1} \in \partial_{\boldsymbol{T}_k} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Gamma}_k^{t+1}, \mathit{\Lambda}^{t+1}\right). \end{array} \right. \tag{91}$$

Proof. Let $\mathcal{X}^{t+1}$ is the optimal solution of $\mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t}, \mathcal{Z}_k^{t}, \boldsymbol{T}_k^{t}, \mathit{\Gamma}_k^{t}, \mathit{\Lambda}^{t}\right)$. For $\mathcal{X}$-subproblem, we have

$$\left\{ \begin{array}{l} 0 \in \partial_{\mathcal{X}} \Phi\left(\mathcal{X}^{t+1}\right) + \mathit{\Lambda}^{t} + \beta\left(\mathcal{X}^{t+1} - \mathcal{C}^{t} \times_1 \boldsymbol{T}_1^{t} \times_2 \cdots \times_d \boldsymbol{T}_d^{t}\right), \\ \mathit{\Lambda}^{t+1} = \mathit{\Lambda}^{t} + \beta\left(\mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \times_2 \cdots \times_d \boldsymbol{T}_d^{t+1}\right), \\ \mathit{\Lambda}^{t+1} = -\partial_{\mathcal{X}} \Phi\left(\mathcal{X}^{t+1}\right). \end{array} \right. \tag{92}$$

Let

$$\boldsymbol{D}_1^{t+1} = \mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t} + \beta\left(\mathcal{C}^{t} \times_1 \boldsymbol{T}_1^{t} \times_2 \cdots \times_d \boldsymbol{T}_d^{t} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \times_2 \cdots \times_d \boldsymbol{T}_d^{t+1}\right), \tag{93}$$

Combining (92) and (93), we can have

$$\begin{aligned} \boldsymbol{D}_1^{t+1} & = \mathit{\Lambda}^{t+1} - \mathit{\Lambda}^{t} + \beta\left(\mathcal{C}^{t} \times_1 \boldsymbol{T}_1^{t} \times_2 \cdots \times_d \boldsymbol{T}_d^{t} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \times_2 \cdots \times_d \boldsymbol{T}_d^{t+1}\right) \\ & + \partial_{\mathcal{X}} \Phi\left(\mathcal{X}^{t+1}\right) + \mathit{\Lambda}^{t} + \beta\left(\mathcal{X}^{t+1} - \mathcal{C}^{t} \times_1 \boldsymbol{T}_1^{t} \times_2 \cdots \times_d \boldsymbol{T}_d^{t}\right) \\ & = \partial_{\mathcal{X}} \Phi\left(\mathcal{X}^{t+1}\right) + \mathit{\Lambda}^{t+1} + \beta\left(\mathcal{X}^{t+1} - \mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \times_2 \cdots \times_d \boldsymbol{T}_d^{t+1}\right) \\ & \in \partial_{\mathcal{X}} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \mathit{\Gamma}_k^{t+1}, \mathit{\Lambda}^{t+1}\right). \end{aligned} \tag{94}$$

Since $\partial_{\mathcal{X}}\Phi\left(\mathcal{X}^{t+1}\right)$ is Lipschitz continuous, we have

$$
\begin{aligned}
\|\boldsymbol{D}_1^{t+1}\|_F^2 &\leq \|\mathit{\Lambda}^{t+1}-\mathit{\Lambda}^t+\beta\left(\mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}\right)\|_F^2 \\
&\leq \|\partial_{\mathcal{X}}\Phi\left(\mathcal{X}^{t+1}\right)-\partial_{\mathcal{X}}\Phi\left(\mathcal{X}^t\right)\|_F^2+\beta^2\|\mathcal{C}^t\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}\|_F^2 \\
&\leq C(\|\mathcal{X}^{t+1}-\mathcal{X}^t\|_F^2)+\beta^2\left(\|\mathcal{C}^{t+1}-\mathcal{C}^t\|_F^2\prod_{k=1}^d\|\boldsymbol{T}_k^t\|_F^2\right) \\
&+\beta^2\left(\|\mathcal{C}^t\|_F^2\sum_{k=1}^d\left(\prod_{p=1}^{k-1}\|\boldsymbol{T}_p^{t+1}\|_F^2\right)(\|\boldsymbol{T}_k^t-\boldsymbol{T}_k^t\|_F^2)\left(\prod_{p=k+1}^d\|\boldsymbol{T}_p^t\|_F^2\right)\right) \\
&\leq C(\beta)(\|\mathcal{X}^{t+1}-\mathcal{X}^t\|_F^2+\|\mathcal{C}^{t+1}-\mathcal{C}^t\|_F^2+\sum_{k=1}^d\|\boldsymbol{T}_k^{t+1}-\boldsymbol{T}_k^t\|_F^2)
\end{aligned}
\tag{95}
$$

where

$$
C\left(\beta\right)\geq\max\left(C,\beta^2\prod_{k=1}^d\|\boldsymbol{T}_k^t\|_F^2,\beta^2\|\mathcal{C}^t\|_F^2\left(\prod_{p=1}^{k-1}\|\boldsymbol{T}_p^{t+1}\|_F^2\right)\left(\prod_{p=k+1}^d\|\boldsymbol{T}_p^t\|_F^2\right)\right).
\tag{96}
$$

Let $\mathcal{C}^{t+1}$ is the optimal solution of $\mathcal{L}\left(\mathcal{X}^{t+1},\mathcal{C}^{t+1},\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\mathit{\Gamma}_k^t,\mathit{\Lambda}^t\right)$. For $\mathcal{C}$-subproblem, we have

$$
\begin{cases}
0\in\sum_{k=1}^d\frac{1}{d}\left\{\mathit{\Gamma}_k^t+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^t\right)\right\}+\mathit{\Lambda}^t+\beta\left(\mathcal{X}^{t+1}-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\right), \\
\mathit{\Gamma}_k^{t+1}=\mathit{\Gamma}_k^t+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^{t+1}\right), \\
\mathit{\Lambda}_k^{t+1}=\mathit{\Lambda}_k^t+\beta\left(\mathcal{X}^{t+1}-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}\right).
\end{cases}
\tag{97}
$$

Let

$$
\begin{aligned}
\boldsymbol{D}_2^{t+1}=&\sum_{k+1}^d\frac{1}{d}\left\{\mathit{\Gamma}_k^{t+1}-\mathit{\Gamma}_k^t+\beta\left(\mathcal{Z}_k^t-\mathcal{Z}_k^{t+1}\right)\right\} \\
&+\mathit{\Lambda}^{t+1}-\mathit{\Lambda}^t+\beta\left(\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}\right).
\end{aligned}
\tag{98}
$$

Combining (97) and (98), we can have

$$
\begin{aligned}
\boldsymbol{D}_2^{t+1}&=\sum_{k+1}^d\frac{1}{d}\left\{\mathit{\Gamma}_k^{t+1}-\mathit{\Gamma}_k^t+\beta\left(\mathcal{Z}_k^t-\mathcal{Z}_k^{t+1}\right)\right\} \\
&+\mathit{\Lambda}^{t+1}-\mathit{\Lambda}^t+\beta\left(\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}\right) \\
&+\sum_{k=1}^d\frac{1}{d}\left\{\mathit{\Gamma}_k^t+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^t\right)\right\}+\mathit{\Lambda}^t+\beta\left(\mathcal{X}^{t+1}-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\right) \\
&=\sum_{k=1}^d\frac{1}{d}\left\{\mathit{\Gamma}_k^{t+1}+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^{t+1}\right)\right\}+\mathit{\Lambda}^{t+1}+\beta\left(\mathcal{X}^{t+1}-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}\right) \\
&\in\partial_{\mathcal{C}}\mathcal{L}\left(\mathcal{X}^{t+1},\mathcal{C}^{t+1},\mathcal{Z}_k^{t+1},\boldsymbol{T}_k^{t+1},\mathit{\Gamma}_k^{t+1},\mathit{\Lambda}^{t+1}\right).
\end{aligned}
\tag{99}
$$

Based on the Cauchy-Schwarz inequality with (98), we can get

$$
\begin{aligned}
\|\boldsymbol{D}_2^{t+1}\|_F^2&\leq\sum_{k+1}^d\frac{1}{d}\left\{\|\mathit{\Gamma}_k^{t+1}-\mathit{\Gamma}_k^t\|_F^2+\beta^2\|\mathcal{Z}_k^{t+1}-\mathcal{Z}_k^t\|_F^2\right\}+\|\mathit{\Lambda}^{t+1}-\mathit{\Lambda}^t\|_F^2 \\
&+\beta^2\|\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^{t+1}\times_2\cdots\times_d\boldsymbol{T}_d^{t+1}-\mathcal{C}^{t+1}\times_1\boldsymbol{T}_1^t\times_2\cdots\times_d\boldsymbol{T}_d^t\|_F^2 \\
&\leq\sum_{k+1}^d\frac{1}{d}\left\{\|\mathit{\Gamma}_k^{t+1}-\mathit{\Gamma}_k^{t+1}\|_F^2+\beta^2\|\mathcal{Z}_k^{t+1}-\mathcal{Z}_k^t\|_F^2\right\}+\|\mathit{\Lambda}^{t+1}-\mathit{\Lambda}^t\|_F^2 \\
&+\beta^2\left(\|\mathcal{C}^{t+1}\|_F^2\sum_{k=1}^d\left(\prod_{p=1}^{k-1}\|\boldsymbol{T}_p^{t+1}\|_F^2\right)(\|\boldsymbol{T}_k^t-\boldsymbol{T}_k^t\|_F^2)\left(\prod_{p=k+1}^d\|\boldsymbol{T}_p^t\|_F^2\right)\right) \\
&\leq 2C\left(\beta\right)\sum_{k=1}^d\frac{1}{d}\left(\|\mathcal{Z}_k^{t+1}-\mathcal{Z}_k^t\|_F^2+\|\boldsymbol{T}_k^{t+1}-\boldsymbol{T}_k^t\|_F^2\right),
\end{aligned}
\tag{100}
$$

where

$$
C\left(\beta\right)\geq\max\left(\beta^2,\beta^2\|\mathcal{C}^{t+1}\|_F^2\left(\prod_{p=1}^{k-1}\|\boldsymbol{T}_p^{t+1}\|_F^2\right)\left(\prod_{p=k+1}^d\|\boldsymbol{T}_p^t\|_F^2\right)\right).
\tag{101}
$$

Let $\mathcal{Z}_k$ is the optimal solution of $\mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^t, \Gamma_k^t, \Lambda^t\right)$. For $\mathcal{Z}_k^{t+1}$-subproblem, we have

$$\left\{\begin{array}{l} 0 \in \partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^{t+1}\right)+\Lambda_k^t+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^{t+1}\right), \\ \Gamma_k^{t+1}=\Gamma_k^t+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^{t+1}\right), \\ \Gamma^{t+1}=-\partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^{t+1}\right). \end{array}\right. \tag{102}$$

Let

$$\boldsymbol{D}_3^{t+1}=\Gamma_k^{t+1}-\Gamma_k^t. \tag{103}$$

Combining(102) and (103), we can have

$$\begin{aligned} \boldsymbol{D}_3^{t+1} & =\Gamma_k^{t+1}-\Gamma_k^t+\partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^{t+1}\right)+\Lambda_k^t+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^{t+1}\right) \\ & =\partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^{t+1}\right)+\Gamma_k^{t+1}+\beta\left(\mathcal{C}_k^{t+1}-\mathcal{Z}_k^{t+1}\right) \\ & \in \partial_{\mathcal{Z}_k} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Gamma_k^{t+1}, \Lambda^{t+1}\right). \end{aligned} \tag{104}$$

Since $\partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^{t+1}\right)$ is Lipschitz continuous, we have

$$\begin{aligned} \|\boldsymbol{D}_3^{t+1}\|_F^2 & \leq\|\Gamma_k^{t+1}-\Gamma_k^t\|_F^2 \\ & =\|\partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^{t+1}\right)-\partial_{\mathcal{Z}_k} \mathcal{L}_3\left(\mathcal{Z}_k^t\right)\|_F^2 \\ & \leq C\|\mathcal{Z}_k^{t+1}-\mathcal{Z}_k^t\|_F^2 \end{aligned} \tag{105}$$

where C is Lipschitz constant.
Let $\boldsymbol{T}_k^{t+1}$ is the optimal solution of $\mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Gamma_k^t, \Lambda^t\right)$. For $\boldsymbol{T}_k$-subproblem, we have

$$\left\{\begin{array}{l} 0 \in \operatorname{Fold}\left(\partial_{\boldsymbol{L}} \Psi\left(\boldsymbol{T}_k^{t+1}\right)-\Lambda^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}+\beta \boldsymbol{X}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\beta \boldsymbol{T}_k^{t+1} \boldsymbol{E}_{(k)}^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}\right), \\ \Lambda^{t+1}=\Lambda^t+\beta\left(\mathcal{X}^{t+1}-\mathcal{C}^t \times_1 \boldsymbol{T}_1^{t+1} \cdots \times_k \boldsymbol{T}_k^{t+1} \times_{k+1} \cdots \times_d^{t+1}\right), \\ \Lambda^{t+1}=-\partial_{\mathcal{X}} \Psi\left(\mathcal{X}^{t+1}\right). \end{array}\right. \tag{106}$$

where $\boldsymbol{E}_{(k)}^t$ is the mode-k matricization of $\mathcal{E}^t=\mathcal{C}^{t+1} \times_1 \boldsymbol{T}_1^{t+1} \cdots \times_{k-1} \boldsymbol{T}_{k-1}^{t+1} \times \boldsymbol{T}_{k+1}^t \cdots \times \boldsymbol{T}_d^t$. Let

$$\boldsymbol{D}_4^{t+1}=\operatorname{Fold}\left(\Lambda^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\Lambda^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}+\beta \boldsymbol{X}_{(k)}^{t+1}\left(\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}\right)+\beta \boldsymbol{T}_k^{t+1}\left(\boldsymbol{E}_{(k)}^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\boldsymbol{E}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}\right)\right) \tag{107}$$

Combining(106) and (107), we can have

$$\begin{aligned} \boldsymbol{D}_4^{t+1}= & \operatorname{Fold}(\Lambda^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\Lambda^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}+\beta \boldsymbol{X}_{(k)}^{t+1}\left(\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}\right)+\partial_{\boldsymbol{L}} \Psi\left(\boldsymbol{T}_k^{t+1}\right)-\Lambda^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top} \\ & +\beta \boldsymbol{T}_k^{t+1}\left(\boldsymbol{E}_{(k)}^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\boldsymbol{E}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}\right)+\beta \boldsymbol{X}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\beta \boldsymbol{T}_k^{t+1} \boldsymbol{E}_{(k)}^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}) \\ = & \partial_{\boldsymbol{L}} \Psi\left(\boldsymbol{T}_k^{t+1}\right)-\Lambda^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}+\beta \boldsymbol{X}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}-\beta \boldsymbol{T}_k^{t+1} \boldsymbol{E}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top} \\ \in & \partial_{\boldsymbol{T}_k} \mathcal{L}\left(\mathcal{X}^{t+1}, \mathcal{C}^{t+1}, \mathcal{Z}_k^{t+1}, \boldsymbol{T}_k^{t+1}, \Gamma_k^{t+1}, \Lambda^{t+1}\right). \end{aligned} \tag{108}$$

Since $\partial_{\boldsymbol{T}_k} \Psi\left(\boldsymbol{T}_k^{t+1}\right)$ is Lipschitz continuous, we have

$$\begin{aligned} & \|\boldsymbol{D}_4^{t+1}\|_F^2=\|Unfold\left(\boldsymbol{D}_4^{t+1}\right)\|_F^2 \\ & =\|\Lambda^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\Lambda^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}+\beta \boldsymbol{X}_{(k)}^{t+1}\left(\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}\right)+\beta \boldsymbol{T}_k^{t+1}\left(\boldsymbol{E}_{(k)}^t\left(\boldsymbol{E}_{(k)}^t\right)^{\top}-\boldsymbol{E}_{(k)}^{t+1}\left(\boldsymbol{E}_{(k)}^{t+1}\right)^{\top}\right)\|_F^2 \\ & \leq 2\|\boldsymbol{E}_{(k)}^t\|\|\Lambda^{t+1}-\Lambda^t\|_F^2+2\beta^2\|\boldsymbol{X}_{(k)}^{t+1}\|_F^2\|\boldsymbol{E}_{(k)}^{t+1}-\boldsymbol{E}_{(k)}^t\|_F^2+2\beta^2\|\boldsymbol{T}_k^{t+1}\|_F^2\|\boldsymbol{E}_{(k)}^{t+1}-\boldsymbol{E}_{(k)}^t\|_F^2 \\ & \leq 2\|\boldsymbol{E}_{(k)}^t\|_F^2\|\partial_T \Psi\left(\boldsymbol{T}_k^{t+1}\right)-\partial_T\left(\boldsymbol{T}_k^t\right)\|_F^2 \\ & +2\beta^2\|\boldsymbol{C}_{(k)}^{t+1}\|_F^2 \sum_{p=1}^{k-1}\|\boldsymbol{T}_p^{t+1}\|_F^2 \sum_{p=k+1}^d\|\boldsymbol{T}_p^t\|_F^2\left(\left(\|\boldsymbol{X}_{(k)}^{t+1}\|_F^2+\|\boldsymbol{T}_k^{t+1}\|_F^2\right)\|\boldsymbol{T}_k^{t+1}-\boldsymbol{T}_k^t\|_F^2\right) \\ & \leq\left(\left(2C+2\beta^2+2\beta^2\|\boldsymbol{X}_{(k)}^{t+1}\|_F^2\right)\|\boldsymbol{C}_{(k)}^{t+1}\| \sum_{p=1}^d\|\boldsymbol{T}_k^{t+1}\|_F^2\right)\|\boldsymbol{T}_k^{t+1}-\boldsymbol{T}_k^t\|_F^2 \\ & \leq 2C(\beta)\|\boldsymbol{T}_k^{t+1}-\boldsymbol{T}_k^{t+1}\|_F^2, \end{aligned} \tag{109}$$

where

$$2C(\beta) \geq\left(2C+2\beta^2+2\beta^2\|\boldsymbol{X}_{(k)}^{t+1}\|_F^2\right)\|\boldsymbol{C}_{(k)}^{t+1}\| \sum_{p=1}^d\|\boldsymbol{T}_k^{t+1}\|_F^2 \tag{110}$$

Based on the above proofs, for any $\boldsymbol{D}_w^{t+1}$, it satisfies the following formulae

$$\|\boldsymbol{D}_w^{t+1}\|_F^2 \leq 2C\left(\beta\right)\left(\|\mathcal{X}^{t+1}-\mathcal{X}^t\|_F^2+\|\mathcal{C}^{t+1}-\mathcal{C}^t\|_F^2+\sum_{k=1}^{d}\frac{1}{d}\left(\|\mathcal{Z}_k^{t+1}-\mathcal{Z}_k^t\|_F^2+\|\boldsymbol{T}_k^{t+1}-\boldsymbol{T}_k^t\|_F^2\right)\right). \tag{111}$$

Therefore, the proof of Lemma 17 is completed. □

Based on the above preparations, we establish the convergence analysis of the proposed algorithm as follows.

Proof. The architecture of proof is to refer to [18], where the proposed algorithm satisfies the following conditions:

1) The sequence $(\mathcal{X}^t,\mathcal{C}^t,\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\Gamma_k^t,\Lambda^t)_{t\in\mathbb{N}}$ satisfies sufficient decrease condition;
2) The sequence $(\mathcal{X}^t,\mathcal{C}^t,\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\Gamma_k^t,\Lambda^t)_{t\in\mathbb{N}}$ is bounded;
3) The $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k,\Lambda)$ is a proper lower semi-continuous function and has the Kurdyka-Łojasiewicz (K-Ł) property at $(\mathcal{X}^t,\mathcal{C}^t,\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\Gamma_k^t,\Lambda^t)_{t\in\mathbb{N}}$;
4) The $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k,\Lambda)$ satisfies subgradient bound condition.

- For condition 1), according to Lemma 15, the sequence $(\mathcal{X}^t,\mathcal{C}^t,\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\Gamma_k^t,\Lambda^t)_{t\in\mathbb{N}}$ satisfies sufficient decrease condition.
- For condition 2), we can obtain that the sequence $(\mathcal{X}^t,\mathcal{C}^t,\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\Gamma_k^t,\Lambda^t)_{t\in\mathbb{N}}$ is bound from the proof of Lemma 16.
- For condition 3), it can verify that $\mathcal{L}_2$ is a C^1 function with locally Lipschitz continuous gradient, and $\mathcal{L}_1,\Phi\left(\mathcal{X}\right)$ and $\Psi\left(\boldsymbol{T}_i\right)$ are proper and lower semi-continuous. Therefore, $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k,\Lambda)$ is a proper lower semi-continuous function. For the K-Ł property, the $\mathcal{L}_2$ is the Frobenius norm, and Frobenius norms are semi-algebraic functions [20]. Since $\Phi\left(\mathcal{X}\right)$ and $\Psi\left(\boldsymbol{T}_i\right)$ are indicator functions with semi-algebraic sets, they are semi-algebraic functions [20]. Then, the nuclear norm term $\mathcal{L}_3(\mathcal{Z}_k)$ is semi-algebraic functions [20]. Thus, the function $L(\mathcal{X},\mathcal{E},\mathcal{Z},\boldsymbol{D},\boldsymbol{T}_i)$ is a semi-algebraic function. Therefore, the function $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k,\Lambda)$ has the K-Ł property at each $(\mathcal{X}^t,\mathcal{C}^t,\mathcal{Z}_k^t,\boldsymbol{T}_k^t,\Gamma_k^t,\Lambda^t)_{t\in\mathbb{N}}$.
- For condition 4), according to Lemma 17, the $\mathcal{L}(\mathcal{X},\mathcal{C},\mathcal{Z}_k,\boldsymbol{T}_k,\Gamma_k,\Lambda)$ satisfies subgradient bound condition.

Therefore, the proof of the convergence analysis is completed. □

APPENDIX H
EXACT RECOVERY GUARANTEE WITH SEMI-ORTHOGONAL MATRIX

Before proving the exact recovery guarantee with semi-orthogonal matrix, we need to discuss the equivalence of elt-product in the original and transformed domains, where the transform is not restricted to the invertible matrix (orthogonal matrix), the non-invertible matrix (semi-orthogonal matrix) is allowed. This bridges the recovery guarantee in the original and transformed domains.

Theorem 6. *For dth-order tensors $\mathcal{A}\in\mathbb{R}^{n_1\times p\times n_3\times\cdots n_d}$ and $\mathcal{B}\in\mathbb{R}^{p\times n_2\times n_3\cdots\times n_d}$ with $\operatorname{rank}(\boldsymbol{A}_{(d)})=\operatorname{rank}(\boldsymbol{B}_{(d)})=\ell_d$ and $0<\ell_d\leq n_d$, for invertible matrix $\boldsymbol{T}_k\in\mathbb{R}^{n_k\times n_k}$, $k\in\{1,2,\cdots,d-1\}$ and full-column rank matrix $\boldsymbol{T}_d\in\mathbb{R}^{n_d\times\ell_d}$, the elt-product is equivalent to*

$$\mathcal{A}\lozenge\mathcal{B}=\Phi^{-1}(\Phi_L(\mathcal{A})\bigtriangleup\Phi_R(\mathcal{B})), \tag{112}$$

where

$$\begin{cases}\Phi^{-1}\left(\mathcal{X}\right)=\mathcal{X}\times_1\left(\boldsymbol{T}_1^\top\right)^{-1}\times_2\left(\boldsymbol{T}_2^\top\right)^{-1}\cdots\times_{d-1}\left(\boldsymbol{T}_{d-1}^\top\right)^{-1}\times_d\left(\boldsymbol{T}_d^\top\right)^\dagger\\ \Phi_L\left(\mathcal{X}\right)=\mathcal{X}\times_1\left(\boldsymbol{T}_1\right)^\top\times_2\left(\boldsymbol{I}_2\right)^\top\cdots\times_{d-1}\left(\boldsymbol{T}_{d-1}\right)^\top\times_d\left(\boldsymbol{T}_d\right)^\top\\ \Phi_R\left(\mathcal{X}\right)=\mathcal{X}\times_1\left(\boldsymbol{I}_1\right)^\top\times_2\left(\boldsymbol{T}_2\right)^\top\cdots\times_{d-1}\left(\boldsymbol{T}_{d-1}\right)^\top\times_d\left(\boldsymbol{T}_d\right)^\top\end{cases} \tag{113}$$

Proof. Let $\boldsymbol{T'}_d=\left[\boldsymbol{T}_d\in\mathbb{R}^{n_d\times\ell_d},0\in\mathbb{R}^{n_d\times(n_d-\ell_d)}\right]$. The SVD of $\boldsymbol{T'}_d$ is $\boldsymbol{T'}_d=[\boldsymbol{U}_1,0]\begin{bmatrix}\boldsymbol{S}_1 & \\ & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^{\mathtt{H}}\\ 0\end{bmatrix}$, and the SVD of $\boldsymbol{T}_d$ is $\boldsymbol{T}_d=\boldsymbol{U}_1\operatorname{diag}(\boldsymbol{S}_1)\boldsymbol{V}_1^{\mathtt{H}}$. Then, the corresponding pseudo-inverse matrices are

$$(\boldsymbol{T'}_d)^\dagger=[\boldsymbol{V}_1,0]\begin{bmatrix}\boldsymbol{S}_1^{-1} & \\ & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_1^{\mathtt{H}}\\ 0\end{bmatrix},(\boldsymbol{T}_d)^\dagger=\boldsymbol{V}_1\mathrm{diag}\left(\boldsymbol{S}_1^{-1}\right)\boldsymbol{U}_1^{\mathtt{H}}. \tag{114}$$

We consider

$$\begin{aligned}&\|\mathcal{A}\times_d\boldsymbol{T'}_d\times_d(\boldsymbol{T'}_d)^\dagger-\mathcal{A}\times_d\boldsymbol{T}_d\times_d\boldsymbol{T}_d^\dagger\|_F^2=\|(\boldsymbol{T'}_d)^\dagger\boldsymbol{T'}_d\boldsymbol{A}_{(d)}-\boldsymbol{T}_d^\dagger\boldsymbol{T}_d\boldsymbol{A}_{(d)}\|_F^2\leq\|(\boldsymbol{T'}_d)^\dagger\boldsymbol{T'}_d-\boldsymbol{T}_d^\dagger\boldsymbol{T}_d\|_F^2\|\boldsymbol{A}_{(d)}\|_F^2\\ =&\|[\boldsymbol{V}_1,0]\begin{bmatrix}\boldsymbol{S}_1^{-1} & \\ & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_1^{\mathtt{H}}\\ 0\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_1 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{S}_1 & \\ & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^{\mathtt{H}}\\ 0\end{bmatrix}-\boldsymbol{V}_1\mathrm{diag}\left(\boldsymbol{S}_1^{-1}\right)\boldsymbol{U}_1^{\mathtt{H}}\boldsymbol{U}_1\boldsymbol{V}_1^{\mathtt{H}}\|_F^2\|\boldsymbol{A}\|_F^2\\ =&\|\begin{bmatrix}\boldsymbol{V}_1 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^{\mathtt{H}}\\ 0\end{bmatrix}-\boldsymbol{V}_1\boldsymbol{V}_1^{\mathtt{H}}\|_F^2\|\boldsymbol{A}\|_F^2=0.\end{aligned} \tag{115}$$

Thus, we have

$$\mathcal{A}\times_d\boldsymbol{T'}_d\times_d(\boldsymbol{T'}_d)^\dagger=\mathcal{A}\times_d\boldsymbol{T}_d\times_d\boldsymbol{T}_d^\dagger. \tag{116}$$

Similarly, we can have

$$\mathcal{B} \times_d \boldsymbol{T'}_d \times_d (\boldsymbol{T'}_d)^{\dagger} = \mathcal{B} \times_d \boldsymbol{T}_d \times_d \boldsymbol{T}_d^{\dagger}. \tag{117}$$

And then, since $\operatorname{rank}(\boldsymbol{A}_{(d)}) = \ell_d$, there is $\operatorname{rank}(\mathcal{A} \times_d \boldsymbol{L}) = \operatorname{rank}(\boldsymbol{L}\boldsymbol{A}_{(d)}) = \ell_d$. Then there exists an invertible matrix $\boldsymbol{L}$ such that $\mathcal{A} \times_d \boldsymbol{L} = \operatorname{Hbfold}\left(\operatorname{diag}\left(\bar{\boldsymbol{A}}_3^{(1)}, \bar{\boldsymbol{A}}_3^{(2)}, \cdots, \bar{\boldsymbol{A}}_3^{(n_3)}, \bar{\boldsymbol{A}}_4^{(1)}, \cdots \bar{\boldsymbol{A}}_4^{(n_4)}, \cdots, \bar{\boldsymbol{A}}_d^{(1)}, \cdots, \bar{\boldsymbol{A}}_d^{(\ell_d)}, 0, \cdots 0\right)\right)$ Therefore, we have

$$\begin{aligned}
\mathcal{A} \times_d \boldsymbol{L} &= \operatorname{Hbfold}\left(\operatorname{diag}\left(\bar{\boldsymbol{A}}_3^{(1)}, \bar{\boldsymbol{A}}_3^{(2)}, \cdots, \bar{\boldsymbol{A}}_3^{(n_3)}, \bar{\boldsymbol{A}}_4^{(1)}, \cdots \bar{\boldsymbol{A}}_4^{(n_4)}, \cdots, \bar{\boldsymbol{A}}_d^{(1)}, \cdots, \bar{\boldsymbol{A}}_d^{(\ell_d)}, 0, \cdots 0\right)\right) \\
&= \operatorname{Fold}\left(\boldsymbol{L}\boldsymbol{A}_{(d)}\right) \\
&= \operatorname{Fold}\left(\begin{bmatrix} \boldsymbol{L}_1 \\ \boldsymbol{L}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_3 & \boldsymbol{U}_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_3 & \\ & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_3^{\mathtt{H}} \\ \boldsymbol{V}_4^{\mathtt{H}} \end{bmatrix}\right) \\
&= \operatorname{Fold}\left(\begin{bmatrix} \boldsymbol{L}_1 \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_3 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_3 & \\ & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_3^{\mathtt{H}} \\ \mathbf{0} \end{bmatrix}\right) \\
&= \operatorname{Fold}\left(\begin{bmatrix} \boldsymbol{L}_1 \\ \mathbf{0} \end{bmatrix} \boldsymbol{A}_d\right) = \mathcal{A} \times_d (\boldsymbol{T'}_d)^{\top}.
\end{aligned} \tag{118}$$

where $\boldsymbol{T'}_d = \left[\boldsymbol{T}_d \in \mathbb{R}^{n_d \times \ell_d}, 0 \in \mathbb{R}^{n_d \times (n_d - \ell_d)}\right]$, $\boldsymbol{T}_d$ is full-column rank. Similarly, we can have $\mathcal{B} \times_d \boldsymbol{L} = \mathcal{B} \times_d (\boldsymbol{T'}_d)^{\top}$. Based on Theorem 1 in our manuscript, for dth-order tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times p \times n_3 \times \cdots n_d}$ and $\mathcal{B} \in \mathbb{R}^{p \times n_2 \times \cdots \times n_{k+1} \times \cdots n_d}$ with invertible matrix $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$, $k \in \{1, 2, \cdots, d-1\}$ and $\boldsymbol{L} \in \mathbb{R}^{n_d \times n_d}$, the elt-product is equivalent to

$$\begin{aligned}
\mathcal{A} \diamond \mathcal{B} = &\left(\left(\mathcal{A} \times_1 \boldsymbol{T}_1^{\top} \times_2 \boldsymbol{I}_2 \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{L}\right) \triangle \left(\mathcal{B} \times_1 \boldsymbol{I}_1 \times_2 \boldsymbol{T}_2^{\top} \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{L}\right)\right) \\
&\times_1 \left(\boldsymbol{T}_1^{\top}\right)^{-1} \times_2 \left(\boldsymbol{T}_2^{\top}\right)^{-1} \cdots \times_{d-1} \left(\boldsymbol{T}_{d-1}^{\top}\right)^{-1} \times_d (\boldsymbol{L})^{-1}.
\end{aligned} \tag{119}$$

In the case where $\mathcal{A}$ and $\mathcal{B}$ have the low-rank property, the transform matrices can be extended from invertible matrices to full-column rank matrices. For dth-order tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times p \times n_3 \times \cdots n_d}$ and $\mathcal{B} \in \mathbb{R}^{p \times n_2 \times n_3 \cdots n_d}$ with $\operatorname{rank}(\boldsymbol{A}_{(d)}) = \operatorname{rank}(\boldsymbol{B}_{(d)}) = \ell_d$, $0 < \ell_d \leq n_d$, and full-column matrix $\boldsymbol{T}_d \in \mathbb{R}^{n_d \times \ell_d}$, the elt-product is equivalent to

$$\begin{aligned}
\mathcal{A} \diamond \mathcal{B} = &\left(\left(\mathcal{A} \times_1 \boldsymbol{T}_1^{\top} \times_2 \boldsymbol{I}_2 \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{L}\right) \triangle \left(\mathcal{B} \times_1 \boldsymbol{I}_1 \times_2 \boldsymbol{T}_2^{\top} \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{L}\right)\right) \\
&\times_1 \left(\boldsymbol{T}_1^{\top}\right)^{-1} \times_2 \left(\boldsymbol{T}_2^{\top}\right)^{-1} \cdots \times_{d-1} \left(\boldsymbol{T}_{d-1}^{\top}\right)^{-1} \times_d (\boldsymbol{L})^{-1} \\
= &\operatorname{Hbfold}\left(\left((\boldsymbol{L})^{-1} \otimes \left(\boldsymbol{T}_{d-1}^{\top}\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_2^{\top}\right)^{-1} \otimes \left(\boldsymbol{T}_1^{\top}\right)^{-1} \otimes \boldsymbol{I}\right)\right. \\
&\operatorname{diag}\left(\bar{\boldsymbol{A}}_3^{(1)}, \bar{\boldsymbol{A}}_3^{(2)}, \cdots, \bar{\boldsymbol{A}}_3^{(n_3)}, \bar{\boldsymbol{A}}_4^{(1)}, \cdots \bar{\boldsymbol{A}}_4^{(n_4)}, \cdots, \bar{\boldsymbol{A}}_d^{(1)}, \cdots, \bar{\boldsymbol{A}}_d^{(\ell_d)}, 0, \cdots 0\right) \\
&\left.\operatorname{diag}\left(\bar{\boldsymbol{B}}_3^{(1)}, \bar{\boldsymbol{B}}_3^{(2)}, \cdots, \bar{\boldsymbol{B}}_3^{(n_3)}, \bar{\boldsymbol{B}}_4^{(1)}, \cdots \bar{\boldsymbol{B}}_4^{(n_4)}, \cdots, \bar{\boldsymbol{B}}_d^{(1)}, \cdots, \bar{\boldsymbol{B}}_d^{(\ell_d)}, 0, \cdots 0\right)\right) \\
= &\operatorname{Hbfold}\left(\left(\begin{bmatrix} \boldsymbol{L}_1^{\dagger} \\ 0 \end{bmatrix} \otimes \left(\boldsymbol{T}_{d-1}^{\top}\right)^{-1} \otimes \cdots \otimes \left(\boldsymbol{T}_2^{\top}\right)^{-1} \otimes \left(\boldsymbol{T}_1^{\top}\right)^{-1} \otimes \boldsymbol{I}\right)\right. \\
&\operatorname{diag}\left(\bar{\boldsymbol{A}}_3^{(1)}, \bar{\boldsymbol{A}}_3^{(2)}, \cdots, \bar{\boldsymbol{A}}_3^{(n_3)}, \bar{\boldsymbol{A}}_4^{(1)}, \cdots \bar{\boldsymbol{A}}_4^{(n_4)}, \cdots, \bar{\boldsymbol{A}}_d^{(1)}, \cdots, \bar{\boldsymbol{A}}_d^{(\ell_d)}, 0, \cdots 0\right) \\
&\left.\operatorname{diag}\left(\bar{\boldsymbol{B}}_3^{(1)}, \bar{\boldsymbol{B}}_3^{(2)}, \cdots, \bar{\boldsymbol{B}}_3^{(n_3)}, \bar{\boldsymbol{B}}_4^{(1)}, \cdots \bar{\boldsymbol{B}}_4^{(n_4)}, \cdots, \bar{\boldsymbol{B}}_d^{(1)}, \cdots, \bar{\boldsymbol{B}}_d^{(\ell_d)}, 0, \cdots 0\right)\right) \\
= &\left(\left(\mathcal{A} \times_1 \boldsymbol{T}_1^{\top} \times_2 \boldsymbol{I}_2 \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{T}_d\right) \triangle \left(\mathcal{B} \times_1 \boldsymbol{I}_1 \times_2 \boldsymbol{T}_2^{\top} \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{T}_d\right)\right) \\
&\times_1 \left(\boldsymbol{T}_1^{\top}\right)^{-1} \times_2 \left(\boldsymbol{T}_2^{\top}\right)^{-1} \cdots \times_{d-1} \left(\boldsymbol{T}_{d-1}^{\top}\right)^{-1} \times_d \begin{bmatrix} \boldsymbol{L}_1^{\dagger} \\ 0 \end{bmatrix} \\
= &\left(\left(\mathcal{A} \times_1 \boldsymbol{T}_1^{\top} \times_2 \boldsymbol{I}_2 \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{T}_d\right) \triangle \left(\mathcal{B} \times_1 \boldsymbol{I}_1 \times_2 \boldsymbol{T}_2^{\top} \cdots \times_{d-1} \boldsymbol{T}_{d-1}^{\top} \times_d \boldsymbol{T}_d\right)\right) \\
&\times_1 \left(\boldsymbol{T}_1^{\top}\right)^{-1} \times_2 \left(\boldsymbol{T}_2^{\top}\right)^{-1} \cdots \times_{d-1} \left(\boldsymbol{T}_{d-1}^{\top}\right)^{-1} \times_d \left(\boldsymbol{T}_d^{\top}\right)^{\dagger} \\
= &\Phi^{-1}(\Phi_L(\mathcal{A}) \triangle \Phi_R(\mathcal{B})).
\end{aligned} \tag{120}$$

□

Remark 1. *In Theorem 1, we consider the low-rankness of $\mathcal{A}$ and $\mathcal{B}$ in mode-d direction. By the similar proof steps, for dth-order tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times p \times n_3 \times \cdots n_d}$ and $\mathcal{B} \in \mathbb{R}^{p \times n_2 \times \cdots \times n_{k+1} \times \cdots n_d}$ with $\operatorname{rank}(\boldsymbol{A}_{(k)}) = \operatorname{rank}(\boldsymbol{B}_{(k)}) = \ell_k$, $0 < \ell_k \leq n_d$, and $\forall k \in [d]$, the transform matrix $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times \ell_k}$ can be extended from invertible matrix to full-column rank matrix, which is beyond the limits of invertible matrix.*

After defining the semi-orthogonal transform-induced elt-product, we can develop a semi-orthogonal element-based tensor singular value decomposition.

Definition 12 (Semi-orthogonal element-based tensor singular value decomposition)**.** *For any dth-order tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $\operatorname{rank}(\boldsymbol{X}_{(d)}) = \ell_d$ *and* $0 < \ell_d \leq n_d$*, semi-orthogonal element-based tensor singular value decomposition (selt-SVD) takes the factorization such that*

$$\mathcal{X} = \mathcal{U} \Diamond \mathcal{S} \Diamond \mathcal{V}^{\mathrm{H}}, \tag{121}$$

where $\mathcal{U}$ *and* $\mathcal{V}$ *are the orthogonal tensors and* $\mathcal{S}$ *is the* f*-diagonal tensor.*

Definition 13 (Enhanced element-based tensor nuclear norm)**.** *For* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with invertible matrix* $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times n_k}$*,* $k \in \{1, 2, \cdots, d-1\}$ *and full-column rank matrix* $\boldsymbol{T}_d \in \mathbb{R}^{n_d \times \ell_d}$*, the enhanced element-based tensor nuclear norm (ETNN$^+$) is defined as*

$$\begin{aligned} \|\mathcal{X}\|_{\mathrm{ETNN}^+} \triangleq \|\mathrm{Hbdiag}\,(\mathcal{X}_\Phi)\|_* &= \sum_{i_3=1}^{n_3} \sum_{i_4=1}^{n_4} \cdots \sum_{i_d=1}^{\ell_d} \|\mathcal{X}_\Phi\,(:,:,i_3,i_4,\cdots,i_d)\|_* \\ &= \sum_{i_3=1}^{n_3} \sum_{i_4=1}^{n_4} \cdots \sum_{i_d=1}^{\ell_d} \|\mathcal{C}\,(:,:,i_3,i_4,\cdots,i_d)\|_*, \end{aligned} \tag{122}$$

where $\mathcal{C} = \mathcal{X}_\Phi$*.*

The following theorem shows that the ETNN$^+$ of a tensor is the convex envelope of the sum of the entries of the transformed rank over a unit ball of the tensor spectral norm.

Theorem 7. *Assuming that dth-order tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $\operatorname{rank}(\boldsymbol{X}_{(k)}) = \ell_k$*,* $0 < \ell_k \leq n_k$*, and* $\forall k \in [d]$*, for any fixed column orthonormal matrix* $\boldsymbol{T}_k \in \mathbb{R}^{n_k \times \ell_k}$*. Let* $\boldsymbol{Q}_k \in \mathbb{R}^{n_k \times (n_k - \ell_k)}$ *be the column complement matrix of* $\boldsymbol{L}_k = [\boldsymbol{T}_k, \boldsymbol{Q}_k]$ *be a orthogonal matrix. Then within the unit ball* $\mathcal{D} = \{\mathcal{X} | \|\mathcal{X}\| \leq 1\}$*, the double conjugate function of* $\operatorname{rank}_{sum}$ *is* $\|\mathcal{X}\|_{\mathrm{ETNN}^+}$*:*

$$\operatorname{rank}_{sum}^{**}(\mathcal{X}) = \|\mathcal{X}\|_{\mathrm{ETNN}^+}. \tag{123}$$

In other words, $\|\mathcal{X}\|_{\mathrm{ETNN}^+}$ *is still the tightest convex envelope of* $\operatorname{rank}_{sum}^{**}(\mathcal{X})$ *within the unit ball* $\mathcal{D}$*.*

Proof. Let $\boldsymbol{Q}_k \in \mathbb{R}^{n_k \times (n_k - \ell_k)}$ be the column complement matrix of $\boldsymbol{L}_k = [\boldsymbol{T}_k, \boldsymbol{Q}_k]$ be a orthogonal matrix. Then we can define an unit ball $\mathcal{D} = \{\mathcal{X} | \|\mathcal{X}\| \leq 1\}$. Furthermore, let $\mathcal{J} = \mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{d-1} \boldsymbol{T}_{d-1}^\top \times_d \boldsymbol{T}_d^\top$ and $\mathcal{H} = \mathcal{Y} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{d-1} \boldsymbol{T}_{d-1}^\top \times_d \boldsymbol{T}_d^\top$, then we have the following:

$$\begin{aligned} &\operatorname{rank}_{\mathrm{sum}}^*(\mathcal{Y}) = \sup_{\mathcal{X} \in \mathcal{D}} \{\langle \mathcal{X}, \mathcal{Y} \rangle - \operatorname{rank}_{\mathrm{sum}}(\mathcal{X})\} \\ &= \sup_{\mathcal{X} \in \mathcal{D}} \Big\{ \Big\langle \mathcal{X} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_d \boldsymbol{T}_d^\top, \mathcal{Y} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_d \boldsymbol{T}_d^\top \Big\rangle - \operatorname{rank}_{\mathrm{sum}}(\mathcal{X}) \\ &\quad + \Big\langle \mathcal{X} \times_1 \boldsymbol{Q}_1^\top \times_2 \boldsymbol{Q}_2^\top \cdots \times_d \boldsymbol{Q}_d^\top, \mathcal{Y} \times_1 \boldsymbol{Q}_1^\top \times_2 \boldsymbol{Q}_2^\top \cdots \times_d \boldsymbol{Q}_d^\top \Big\rangle \Big\} \\ &= \sum_{i_3=1}^{\ell_3} \sum_{i_4=1}^{\ell_4} \cdots \sum_{i_d=1}^{\ell_d} rank\,(\mathcal{H}\,(:,:,i_3,i_4,\cdots,i_d)) \\ &\quad + \sum_{i_3=\ell_3+1}^{n_3} \sum_{i_4=\ell_4+1}^{n_4} \cdots \sum_{i_d=\ell_d+1}^{n_d} \sup_{\|\mathcal{J}\| \leq 1} \langle \mathcal{J}\,(:,:,i_3,i_4,\cdots,i_d), \mathcal{H}\,(:,:,i_3,i_4,\cdots,i_d) \rangle \\ &= \sum_{i_3=1}^{\ell_3} \sum_{i_4=1}^{\ell_4} \cdots \sum_{i_d=1}^{\ell_d} (\sigma\,(\mathcal{H}\,(:,:,i_3,i_4,\cdots,i_d)) - 1)_+ + \sum_{i_3=\ell_3+1}^{n_3} \sum_{i_4=\ell_4+1}^{n_4} \cdots \sum_{i_d=\ell_d+1}^{n_d} \sigma\,(\mathcal{H}\,(:,:,i_3,i_4,\cdots,i_d)) \end{aligned} \tag{124}$$

Then, the double conjugate function is defined as follows:

$$\begin{aligned} &\operatorname{rank}_{\mathrm{sum}}^{**}(\mathcal{Z}) = \sup_{\mathcal{Y}} \{\langle \mathcal{Y}, \mathcal{Z} \rangle - \operatorname{rank}_{\mathrm{sum}}^*(\mathcal{Y})\} \\ &= \sup_{\mathcal{Y}} \Bigg\{ \langle \mathcal{Y}, \mathcal{Z} \rangle - \sum_{i_3=1}^{\ell_3} \sum_{i_4=1}^{\ell_4} \cdots \sum_{i_d=1}^{\ell_d} (\sigma\,(\mathcal{Y}\,(:,:,i_3,i_4,\cdots,i_d)) - 1)_+ + \sum_{i_3=\ell_3+1}^{n_3} \sum_{i_4=\ell_4+1}^{n_4} \cdots \sum_{i_d=\ell_d+1}^{n_d} \sigma\,(\mathcal{Y}\,(:,:,i_3,i_4,\cdots,i_d)) \Bigg\} \\ &= \sup_{\mathcal{Y}} \Bigg\{ \sum_{i_3=1}^{\ell_3} \sum_{i_4=1}^{\ell_4} \cdots \sum_{i_d=1}^{\ell_d} \sigma\,(\mathcal{Y}\,(:,:,i_3,i_4,\cdots,i_d))\,\sigma\,(\mathcal{Z}\,(:,:,i_3,i_4,\cdots,i_d)) - (\sigma\,(\mathcal{Y}\,(:,:,i_3,i_4,\cdots,i_d)) - 1)_+ \\ &\quad + \sum_{i_3=\ell_3+1}^{n_3} \sum_{i_4=\ell_4+1}^{n_4} \cdots \sum_{i_d=\ell_d+1}^{n_d} \sigma\,(\mathcal{Y}\,(:,:,i_3,i_4,\cdots,i_d))\,\sigma\,(\mathcal{Z}\,(:,:,i_3,i_4,\cdots,i_d)) - \sigma\,(\mathcal{Y}\,(:,:,i_3,i_4,\cdots,i_d)) \Bigg\}. \end{aligned} \tag{125}$$

Therefore, $\|\mathcal{Z} \times_1 \boldsymbol{T}_1^\top \times_2 \boldsymbol{T}_2^\top \cdots \times_{d-1} \boldsymbol{T}_{d-1}^\top \times_d \boldsymbol{T}_d^\top\| \leq 1$ and $\|\mathcal{Z} \times_1 \boldsymbol{L}_1^\top \times_2 \boldsymbol{L}_2^\top \cdots \times_{d-1} \boldsymbol{L}_{d-1}^\top \times_d \boldsymbol{L}_d^\top\| \leq 1$ hold, which implies that $\|\mathcal{Z} \times_1 \boldsymbol{L}_1^\top \times_2 \boldsymbol{L}_2^\top \cdots \times_{d-1} \boldsymbol{L}_{d-1}^\top \times_d \boldsymbol{L}_d^\top\| \leq 1$. In this case, for $i_k = 1, \cdots, \ell_k$ and $\forall k \in \{3, \cdots, d\}$, we have $\sigma(\mathcal{Y}(:,:,i_3,i_4,\cdots,i_d)) = 1$, and for $i_k = \ell_k, \cdots, n_k$ and $\forall k \in \{3, \cdots, d\}$, we have $\sigma(\mathcal{Y}(:,:,i_3,i_4,\cdots,i_d)) = 0$. Then we have the following result:

$$\operatorname{rank}_{\text{sum}}^{**}(\mathcal{Z}) = \sum_{i_3=1}^{\ell_3}\sum_{i_4=1}^{\ell_4}\cdots\sum_{i_d=1}^{\ell_d} \sigma\left(\mathcal{Z}(:,:,i_3,i_4,\cdots,i_d)\right) = \|\mathcal{Z}\|_{\text{ETNN}^+}. \tag{126}$$

$\square$

Based on the low-rank metric, we propose the LRTC model via ETNN$^+$, the model can be expressed as

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\text{ETNN}^+}, \text{ s.t. } \mathcal{X}_\Omega = \mathcal{O}_\Omega. \tag{127}$$

With the aforementioned ETNN$^+$, we examine the exact recovery guarantee for the proposed model (127). Before establishing our main result, we first need to define the tensor incoherence conditions.

Definition 14 (Tensor incoherence conditions)**.** *For $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ with $\operatorname{rank}(\boldsymbol{X}_{(d)}) = \ell_d$ and $0 < \ell_d \leq n_d$, suppose that the selt-SVD is $\mathcal{X} = \mathcal{U}\lozenge\mathcal{S}\lozenge\mathcal{V}^{\mathtt{H}}$ with transformed multi-rank $\boldsymbol{R}$. Then, the tensor incoherence conditions with parameter $\mu > 1$ are given by*

$$\max_{i_k=1,\cdots,n_k,k\neq 2} \|\mathcal{U}^{\mathtt{H}} \diamond \mathring{e}_1^{(i_1)}\|_F^2 \leq \frac{\mu\sum_{i=1}^L R(i)}{n_1 L} \tag{128}$$

$$\max_{i_k=1,\cdots,n_k,k\neq 1} \|\mathcal{V}^{\mathtt{H}} \diamond \mathring{e}_2^{(i_2)}\|_F^2 \leq \frac{\mu\sum_{i=1}^L R(i)}{n_2 L} \tag{129}$$

where $\mathring{e}_1^{(i_1)}$ is the standard dth-order tensor basis whose size is $n_1 \times 1 \times n_3 \cdots \times n_d$ with its $(i_1, 1, i_3 \cdots, i_d)$-th entry be 1 and be 0 otherwise, the $\mathring{e}_2^{(i_2)} = (\mathring{e}_1^{(i_1)})^{\mathtt{H}}$, and $L = n_3 n_4 \cdots \ell_d$.

Based on the tensor incoherence conditions (128) and (129), we can have the following Theorem.

Theorem 8. *For $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ with $\operatorname{rank}(\boldsymbol{X}_{(d)}) = \ell_d$ and $0 < \ell_d \leq n_d$, suppose that the selt-SVD is $\mathcal{X} = \mathcal{U}\lozenge\mathcal{S}\lozenge\mathcal{V}^{\mathtt{H}}$ with transformed multi-rank $\boldsymbol{R}$, they are satisfied with $L = n_3 n_4 \cdots \ell_d$, $n^{(1)} = \max(n_1, n_2)$, $n^{(2)} = \min(n_1, n_2)$, and $\dot{n} = n_1^{(1)} L$. Suppose that the indices set $\Omega \backsim Ber(\rho)$ with $|\Omega| = m$ and the tensor incoherence conditions (128)-(129) hold. Then, there exist universal constants $c_1, c_2, c_3 > 0$ such that $\mathcal{X}$ is the unique solution with probability at least $1 - c_1 \dot{n}^{c_2}$, provided that*

$$m \geq c_3 \mu \sum_{i=1}^{L} R(i)\, n^{(1)} \log\left(n^{(1)} L\right). \tag{130}$$

Remark 2. *For the elt-product, it depends on the transform. In the following proofs, we focus on Φ with orthogonal matrix $\boldsymbol{T}_k \in \mathbb{R}^{n_k\times n_k}$, $k \in \{1, 2, \cdots, d-1\}$ and semi-orthogonal matrix $\boldsymbol{T}_d \in \mathbb{R}^{n_d\times \ell_d}, 0 < \ell_d \leq n_d$.*

A. Main Preliminaries

In this subsection, we introduce main preliminaries and useful lemmas, which play an important role in our proof.

Definition 15. *The $\ell_{\infty,2}$-norm of the tensor $\mathcal{X}$ is defined as*

$$\|\mathcal{X}\|_{\infty,2} = \max\{\max_{(i_1,i_3,\cdots,i_d)} \|\mathcal{X}(i_1,:,i_3,\cdots,i_d)\|_F, \max_{(i_2,i_3,\cdots,i_d)} \|\mathcal{X}(:,i_2,i_3,\cdots,i_d)\|_F\}. \tag{131}$$

Definition 16. *For $i_1 \in [n_1], \cdots, i_d \in [n_d]$, we define*

$$e_{i_1\cdots i_d} := \left(\mathring{e}_1^{(i_1)}\right) \lozenge \left(\dot{e}_3^{(i_3)}\right) \left(\dot{e}_4^{(i_4)}\right) \lozenge \cdots \lozenge \left(\dot{e}_d^{(i_d)}\right) \lozenge \left(\mathring{e}_2^{(i_2)}\right), \tag{132}$$

where an $n_1 \times \cdots \times n_d$ sized tensor with its $(i_1, \cdots, i_d)$ -th entry equaling to 1 and the rest equaling to 0; $\mathring{e}_k^{(i_k)}$ is the standard dth-order tensor basis whose size is $n_1 \times \cdots \times n_k \times 1 \times n_{k+2} \times \cdots \times n_d$ with its $(i_1, \cdots, i_k, 1, i_{k+2}, \cdots, i_d)$-th entry equaling to 1 and the rest equaling to 0, and the $\mathring{e}_{k+1}^{(i_{k+1})} = (\mathring{e}_k^{(i_k)})^{\top_k}$; $\dot{e}_{k+2}^{(i_{k+2})}$ is a tensor with the $(i_1, \cdots, i_k, 1, 1, i_{k+2}, \cdots, i_d)$-th entry of $\Phi(\dot{e}_{k+2}^{(i_{k+2})})$ equaling to $((\Phi(\vec{e}_{k+2}^{(i_{k+2})}))_{i_1,\cdots,i_k,1,1,i_{k+2},\cdots,i_d})^{-1}$ if $((\Phi(\vec{e}_{k+2}^{(i_{k+2})}))_{i_1,\cdots,i_k,1,1,i_{k+2},\cdots,i_d})^{-1} \neq 0$ and 0 otherwise.

Given any dth-order tensor $\mathcal{X}$, we have

$$\begin{aligned} \mathcal{X} &= \sum_{i_1\cdots i_d} \langle e_{i_1\cdots i_d}, \mathcal{X}\rangle e_{i_1\cdots i_d}, \\ &= \sum_{i_1\cdots i_d} \mathcal{X}_{,i_1\cdots i_d} \left(\mathring{e}_1^{(i_1)}\right) \lozenge \left(\dot{e}_3^{(i_3)}\right) \left(\dot{e}_4^{(i_4)}\right) \lozenge \cdots \lozenge \left(\dot{e}_d^{(i_d)}\right) \lozenge \left(\mathring{e}_2^{(i_2)}\right). \end{aligned} \tag{133}$$

Definition 17. *For some $\mathcal{A}$ and $\mathcal{B}$, we define $\mathbb{T}$ to be the linear space*

$$\mathbb{T} := \left\{ \mathcal{W} \left| \mathcal{W} = \mathcal{U} \lozenge \mathcal{A}^{\mathtt{H}} + \mathcal{B} \lozenge \mathcal{V}^{\mathtt{H}} \right. \right\}, \tag{134}$$

and $\mathbb{T}^{\perp}$ to be the orthogonal complement of $\mathbb{T}$. The orthogonal projection $\mathcal{P}_{\mathbb{T}}$ on $\mathbb{T}$ is given by

$$\mathcal{P}_{\mathbb{T}} \left(\mathcal{Z} \right) = \mathcal{U} \lozenge \mathcal{U}^{\mathtt{H}} \lozenge \mathcal{Z} + \mathcal{Z} \lozenge \mathcal{V} \lozenge \mathcal{V}^{\mathtt{H}} - \mathcal{U} \lozenge \mathcal{U}^{\mathtt{H}} \lozenge \mathcal{Z} \lozenge \mathcal{V} \lozenge \mathcal{V}^{\mathtt{H}}, \tag{135}$$

and $\mathcal{P}_{\mathbb{T}^{\perp}}$ is defined as

$$\mathcal{P}_{\mathbb{T}^{\perp}} \left(\mathcal{Z} \right) = \mathcal{Z} - \mathcal{P}_{\mathbb{T}} \left(\mathcal{Z} \right) = \left(\mathcal{I} - \mathcal{U} \lozenge \mathcal{U}^{\mathtt{H}} \right) \lozenge \mathcal{Z} \lozenge \left(\mathcal{I} - \mathcal{V} \lozenge \mathcal{V}^{\mathtt{H}} \right). \tag{136}$$

Lemma 18. *For any tensor $\mathcal{X}$ with transformed multi-rank $\boldsymbol{R}$, and $\mathbb{T}$ be given as* (134). *Suppose that the tensor incoherence conditions* (128) *and* (129) *are satisfied, then we have*

$$\left\| \mathcal{P}_{\mathbb{T}} \left(e_{i_1 \cdots i_d} \right) \right\|_F^2 \leq \frac{2\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}{n_1^{(1)} (L)^2}. \tag{137}$$

Lemma 19. *For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $\operatorname{rank}(\boldsymbol{X}_{(d)}) = \ell_d$ and $0 < \ell_d \leq n_d$, suppose that the selt-SVD is $\mathcal{X} = \mathcal{U} \lozenge \mathcal{S} \lozenge \mathcal{V}^{\mathtt{H}}$. Then the subdifferential (the set of subgradients) of $\| \cdot \|_{\mathrm{ETNN}^+}$ at $\mathcal{X}$ can be described as:*

$$\partial \| \mathcal{X} \|_{\mathrm{ETNN}^+} = \left\{ \mathcal{U} \lozenge \mathcal{V}^{\mathtt{H}} + \mathcal{W} \right| \mathcal{U}^{\mathtt{H}} \lozenge \mathcal{W} = 0, \mathcal{W} \lozenge \mathcal{V} = 0, \| \mathcal{W} \| \leq 1 \left. \right\}. \tag{138}$$

Lemma 20. *Suppose that Ω is sampled from the Bernoulli model with parameter $\rho = \frac{m}{\prod_{i=1}^{d} n_i}$. Let $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \cdots, \times n_d}$ then with the high probability*

$$\left\| \rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \right\| \leq \epsilon, \tag{139}$$

and

$$\left\| \left(\rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \right) \mathcal{Z} \right\|_{\infty} \leq \epsilon \left\| \mathcal{P}_{\mathbb{T}} \mathcal{Z} \right\|_{\infty}, \tag{140}$$

provided that $\rho \geq c_5 \epsilon^{-2} \frac{\mu \sum_{i=1}^{L} \boldsymbol{R}(i) \log n_1^{(1)} L}{n_1^{(2)} L}$ for some positive numerical constant $c_5 > 0$.

Lemma 21. *Suppose that $\mathcal{Z}$ is fixed, and $\Omega \sim Ber(\rho)$. Then with high probability,*

$$\left\| \left(\rho^{-1} \mathcal{P}_{\Omega} - \mathcal{I} \right) \mathcal{Z} \right\| \leq c \left(\frac{\log(n_1^{(1)} L)}{\rho} \| \mathcal{Z} \|_{\infty} + \sqrt{\frac{\log(n_1^{(1)} L)}{\rho}} \| \mathcal{Z} \|_{\infty, 2} \right), \tag{141}$$

provided that $\rho \geq c_4 \epsilon^{-2} \frac{\mu \sum_{i=1}^{L} \boldsymbol{R}(i) \log n_1^{(1)} L}{n_1^{(2)} L}$ for some positive numerical constant $c_4 > 0$.

Lemma 22. *Suppose that $\mathcal{Z}$ is fixed, and $\Omega \sim Ber(\rho)$. Then with high probability,*

$$\| \rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Omega} \mathcal{Z} - \mathcal{P}_{\mathbb{T}} \mathcal{Z} \|_{\infty, 2} \leq \frac{1}{2} \sqrt{\frac{n_1^{(1)} L}{\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}} \| \mathcal{Z} \|_{\infty} + \frac{1}{2} \| \mathcal{Z} \|_{\infty, 2}, \tag{142}$$

provided that $\rho \geq c_3 \epsilon^{-2} \frac{\mu \sum_{i=1}^{L} \boldsymbol{R}(i) \log n_1^{(1)} L}{n_1^{(2)} L}$ for some positive numerical constant $c_3 > 0$.

Lemma 23. *Assume that $\left\| \rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \right\| \leq \frac{1}{2}$. Then for all $\mathcal{Z}$, we have*

$$\| \mathcal{P}_{\mathbb{T}} (\mathcal{Z}) \|_F \leq \sqrt{\frac{2}{\rho}} \| \mathcal{P}_{\mathbb{T}^{\perp}} (\mathcal{Z}) \|_{\mathrm{ETNN}^+}, \forall \mathcal{Z} \in \left\{ \mathcal{Z}' : \mathcal{P}_{\Omega} \left(\mathcal{Z}' \right) = 0 \right\}. \tag{143}$$

Lemma 24. *If there exists $\mathbb{T}$ such that*

$$\left\| \rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Omega} \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \right\| \leq \frac{1}{2}, \tag{144}$$

and a tensor $\mathcal{Y}$ satisfying

$$\| \mathcal{P}_{\mathbb{T}} \left(\mathcal{Y} - \mathcal{T} \right) \|_F \leq \frac{1}{4 n_1^{(1)} L}, \tag{145}$$

$$\| \mathcal{P}_{\mathbb{T}^{\perp}} \left(\mathcal{Y} \right) \|_F \leq \frac{1}{2}, \tag{146}$$

$$\mathcal{P}_{\Omega^{\perp}} \left(\mathcal{Y} \right) = 0. \tag{147}$$

Then $\mathcal{X}_0$ is the unique solution when $n_1^{(1)} L$ for any $k \in [d]$ is sufficiently large.

B. Proof of Theorem 8

Proof. Consider a feasible perturbation $\mathcal{X}_0 + \triangle$, where $\mathcal{P}_{\Omega^\perp}(\triangle) = \triangle$. We now show that the objective value $f(\mathcal{X}_0 + \triangle)$ is strictly greater than $f(\mathcal{X}_0)$ unless $\triangle = 0$. Due to

$$\mathcal{U}_0 \Diamond \mathcal{V}_0^{\mathtt{H}} + \mathcal{W}_0 \in \partial \|\mathcal{X}_0\|_{\mathrm{ETNN}^+},$$

we can get

$$\mathcal{P}_{\mathbb{T}}(\mathcal{W}_0) = 0, \|\mathcal{W}_0\| \leq 1.$$

We have

$$\begin{aligned}
& f\left(\mathcal{X}_0 + \triangle\right) - f\left(\mathcal{X}_0\right) \\
& \geq \left\langle \mathcal{U}_0 \diamond \mathcal{V}_0^H + \mathcal{W}_0, \triangle \right\rangle \\
& = \left\langle \mathcal{T}_0 + \mathcal{W}_0, \triangle \right\rangle \\
& = \left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} + \left\langle \mathcal{T}_0, \triangle \right\rangle \\
& = \left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} + \left\langle \mathcal{T}_0, \triangle \right\rangle - \left\langle \mathcal{Y}, \triangle \right\rangle && (148) \\
& = \left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} + \left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\left(\triangle\right) \right\rangle - \left\langle \mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{Y}\right), \mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right) \right\rangle \\
& \geq \left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} - \left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\left(\triangle\right) \right\rangle - \left\langle \mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{Y}\right), \mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right) \right\rangle && (149) \\
& \geq \left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} - \frac{1}{4n_1^{(1)}L}\left\|\mathcal{P}_{\mathbb{T}}\left(\triangle\right)\right\|_F - \frac{1}{2}\left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} && (150) \\
& = \frac{1}{2}\left\|\mathcal{P}_{\mathbb{T}^\perp}\left(\triangle\right)\right\|_{\mathrm{ETNN}^+} - \frac{1}{4n_1^{(1)}L}\left\|\mathcal{P}_{\mathbb{T}}\left(\triangle\right)\right\|_F \\
& \geq \left(\frac{1}{2}\sqrt{\frac{\rho}{2}} - \frac{1}{4n_1^{(1)}L}\right)\left\|\mathcal{P}_{\mathbb{T}}\left(\triangle\right)\right\|_F && (151)
\end{aligned}$$

When $n_1^{(1)}L$ is large enough such that

$$\frac{1}{2}\sqrt{\frac{\rho}{2}} - \frac{1}{4n_1^{(1)}L} > 0.$$

Now, we give the proof of (148) - (150). For (148) , since $\mathcal{P}_{\Omega^\perp}\mathcal{Y} = 0$ and $\mathcal{P}_{\Omega^\perp}(\triangle) = \triangle$, we have

$$\begin{aligned}
& \langle \mathcal{Y}, \triangle \rangle \\
= & \left(\left\langle \left(\mathcal{P}_\Omega + \mathcal{P}_{\Omega^\perp}\right)\left(\mathcal{Y}\right), \left(\mathcal{P}_\Omega + \mathcal{P}_{\Omega^\perp}\right)\left(\triangle\right) \right\rangle\right) \\
= & \left(\langle \mathcal{P}_\Omega\mathcal{Y}, \mathcal{P}_\Omega\triangle \rangle + \langle \mathcal{P}_\Omega\mathcal{Y}, \mathcal{P}_{\Omega^\perp}\triangle \rangle\right) + \left(\langle \mathcal{P}_{\Omega^\perp}\mathcal{Y}, \mathcal{P}_\Omega\triangle \rangle + \langle \mathcal{P}_{\Omega^\perp}\mathcal{Y}, \mathcal{P}_{\Omega^\perp}\triangle \rangle\right) \\
= & 0.
\end{aligned}$$

For (149), we have

$$\begin{aligned}
& \langle \mathcal{T}_0 - \mathcal{Y}, \triangle \rangle \\
& = \left(\left\langle \left(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\mathbb{T}^\perp}\right)\left(\mathcal{T}_0 - \mathcal{Y}\right), \left(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\mathbb{T}^\perp}\right)\left(\triangle\right) \right\rangle\right) \\
& = \left(\left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\left(\triangle\right) \right\rangle + \left\langle \mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\triangle \right\rangle\right) \\
& + \left(\left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}^\perp}\triangle \right\rangle + \mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}^\perp}\triangle\right) \\
& = \left(\left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\triangle \right\rangle + \left\langle \mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}^\perp}\triangle \right\rangle\right) \\
& = \left(\left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\triangle \right\rangle + \left\langle \mathcal{P}_{\mathbb{T}^\perp}\mathcal{T}_0, \mathcal{P}_{\mathbb{T}^\perp}\triangle \right\rangle - \left\langle \mathcal{P}_{\mathbb{T}^\perp}\mathcal{Y}, \mathcal{P}_{\mathbb{T}^\perp}\triangle \right\rangle\right) \\
& = \left(\left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\triangle \right\rangle - \left\langle \mathcal{P}_{\mathbb{T}^\perp}\mathcal{Y}, \mathcal{P}_{\mathbb{T}^\perp}\triangle \right\rangle\right).
\end{aligned}$$

For (150), according to (145) and (146) , we have

$$\left\langle \mathcal{P}_{\mathbb{T}}\left(\mathcal{T}_0 - \mathcal{Y}\right), \mathcal{P}_{\mathbb{T}}\triangle \right\rangle \geq \frac{1}{4n_1^{(1)}L}\|\mathcal{P}_{\mathbb{T}}\triangle\|_F,$$

$$\left\langle \mathcal{P}_{\mathbb{T}^\perp}\mathcal{Y}, \mathcal{P}_{\mathbb{T}^\perp}\triangle \right\rangle \geq \|\mathcal{P}_{\mathbb{T}^\perp}\triangle\|_{\mathrm{ETNN}^+}.$$

$\square$

The above proof assumes that the following three conditions hold, i.e.,

$$\|\mathcal{P}_{\mathbb{T}}\left(\mathcal{Y}-\mathcal{T}\right)\|_F \leq \frac{1}{4n_1^{(1)}L}, \tag{152}$$

$$\|\mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{Y}\right)\|_F \leq \frac{1}{2}, \tag{153}$$

$$\mathcal{P}_{\Omega^\perp}\left(\mathcal{Y}\right) = 0. \tag{154}$$

We apply the Golfing Scheme [13] to prove these three conditions, where the architecture of proof is to refer to [3, 14].

Proof. □

We will use an approach called Golfing Scheme introduced by Gross [15] and we will follow the idea in [16, 17] where the strategy is to construct $\mathcal{Y}$ iteratively. Let Ω be a union of smaller sets Ω^t such that $\Omega = \bigcup_{t=1}^{t_0} \Omega^t$ where $t_0 = \lfloor 5\log\left(n_1^{(1)}L\right)\rfloor$. For each t , we assume $q := 1-(1-p)^{1/t}$, and it is easy to verify that it's equivalent to our original Ω. Let $\mathcal{W}^0 = 0$ and for $t = 1, 2, , \cdots t_0$,

$$\mathcal{W}^t = \mathcal{W}^{t-1} + q^{-1}\mathcal{P}_{\Omega^t}\mathcal{P}_{\mathbb{T}}\left(\mathcal{T} - \mathcal{P}_{\mathbb{T}}\left(\mathcal{W}^{t-1}\right)\right), \tag{155}$$

and $\mathcal{Y} = \mathcal{W}_{t_0}$. By this construction we can get $\mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{Y}$.

For $t = 0, 1, \ldots, t_0$, set $\mathcal{D}^t = \mathcal{T} - \mathcal{P}_T\left(\mathcal{W}_t\right)$. Then we have $\mathcal{D}^0_{\langle k\rangle} = \mathcal{T}$ and

$$\mathcal{D}^t = \left(\mathcal{P}_{\mathbb{T}} - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^t}\mathcal{P}_{\mathbb{T}}\right)\left(\mathcal{D}^{t-1}\right).$$

Note that Ω^t is independent of $\mathcal{D}^t$, which implies

$$\left\|\mathcal{D}^t\right\|_F \leq \left\|\mathcal{P}_{\mathbb{T}} - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^t}\mathcal{P}_{\mathbb{T}}\right\|\left\|\mathcal{D}^{t-1}\right\|_F \leq \frac{1}{2}\left\|\mathcal{D}^{t-1}\right\|_F.$$

Since $q \geq p/t_0 \geq c_5\mu\sum_{i=1}^L \boldsymbol{R}(i)\log\left(n_1^{(1)}L\right)/n$, we have

$$\begin{aligned}
&\|\mathcal{P}_T(\mathcal{Y}) - \mathcal{T}\|_F = \left\|\mathcal{D}^{t_0}\right\|_F \\
&\leq \left(\frac{1}{2}\right)^{t_0}\|\mathcal{T}\|_F \leq \frac{1}{2\left(n_1^{(1)}L\right)^2}\sqrt{\sum_{i=1}^L \boldsymbol{R}(i)} \leq \frac{1}{2n_1^{(1)}L^2},
\end{aligned}$$

holds with high probability for some large enough constants $c_5 > 0$.

From (155) we know that $\mathcal{Y} = \mathcal{W}^{t_0} = \sum_{t=1}^{t_0}\left(q^{-1}\mathcal{P}_{\Omega^t}\mathcal{P}_T\right)\left(\mathcal{D}^{t-1}\right)$, so use Lemma 21 we obtain for some constant $c_4 > 0$,

$$\begin{aligned}
&\|\mathcal{P}_{\mathbb{T}^\perp}\left(\mathcal{Y}\right)\|_F \\
&\leq \sum_{t=1}^{t_0}\|\mathcal{P}_{\mathbb{T}^\perp}q^{-1}\mathcal{P}_{\Omega^t}\mathcal{P}_{\mathbb{T}}\left(\mathcal{D}^t\right)\|_F \\
&\leq \sum_{t=1}^{t_0}\|\left(q^{-1}\mathcal{P}_{\Omega^t} - \mathcal{I}\right)\mathcal{P}_{\mathbb{T}}\left(\mathcal{D}^{t-1}\right)\|_F \\
&\leq c_4\sum_{t=1}^{t_0}\left(\frac{\log\left(n_1^{(1)}L\right)}{m}\|\mathcal{D}^{t-1}\|_\infty + \sqrt{\frac{\log\left(n_1^{(1)}L\right)}{m}}\|\mathcal{D}^{t-1}\|_{\infty,2}\right) \\
&\leq \frac{c_4}{c_3}\sum_{t=1}^{t_0}\left(\frac{n_1^{(1)}L}{\mu\sum_{i=1}^L \boldsymbol{R}(i)}\|\mathcal{D}^{t-1}\|_\infty + \sqrt{\frac{n_1^{(1)}L}{\mu\sum_{i=1}^L \boldsymbol{R}(i)}}\|\mathcal{D}^{t-1}\|_{\infty,2}\right),
\end{aligned} \tag{156}$$

where we can bound term $\left\|\mathcal{D}^{t-1}\right\|_\infty$ using Lemma 20 as follows,

$$\begin{aligned}
\|\mathcal{D}^{t-1}\|_\infty &\leq \|\mathcal{P}_{\mathbb{T}} - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^t}\mathcal{P}_{\mathbb{T}}\|\|\mathcal{D}^{t-1}\|_\infty \leq \frac{1}{2}\|\mathcal{D}^{t-2}\|_\infty \\
&\leq \frac{1}{2}\|\mathcal{P}_{\mathbb{T}} - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^t}\mathcal{P}_{\mathbb{T}}\|\|\mathcal{D}^{t-3}\|_\infty \leq \left(\frac{1}{2}\right)^2\|\mathcal{D}^{t-3}\|_\infty \\
&\leq \cdots \leq \left(\frac{1}{2}\right)^{t_0-1}\|\mathcal{D}^0\|_\infty = \left(\frac{1}{2}\right)^{t_0-1}\|\mathcal{T}^0\|_\infty.
\end{aligned} \tag{157}$$

Based on Lemma 22, combining (156) and (157), we have

$$
\begin{aligned}
&\|\mathcal{P}_{T^{\perp}}(\mathcal{Y})\|_F \\
\leq &\frac{c_4}{\sqrt{c_3}} \frac{n_1^{(1)} L}{\mu \sum_{i=1}^{L} \boldsymbol{R}(i)} \|\mathcal{T}\|_{\infty} \sum_{t=1}^{t_0} (t+1) \left(\frac{1}{2}\right)^{t-1} + \frac{c_4}{\sqrt{c_3}} \sqrt{\frac{n_1^{(1)} L}{\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}} \|\mathcal{T}\|_{\infty,2} \sum_{t=1}^{t_0} \left(\frac{1}{2}\right)^{t-1} \\
\leq &\frac{6c_4}{\sqrt{c_3}} \frac{n_1^{(1)} L}{\mu \sum_{i=1}^{L} \boldsymbol{R}(i)} \|\mathcal{T}\|_{\infty} + \frac{2c_4}{\sqrt{c_3}} \sqrt{\frac{n_1^{(1)} L}{\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}} \|\mathcal{T}\|_{\infty,2},
\end{aligned}
$$

holds with high probability for some large enough constants $c_4, c_3 > 0$. On the other hand, we have

$$
\begin{aligned}
&\|\mathcal{T}^0_{\langle k\rangle}\|_{\infty,2} = \max\{ \max_{(i_1,\cdots,i_k,i_{k+3},\cdots,i_d)} \|\mathcal{T}(i_1,\cdots,i_k,:,:,i_{k+3},\cdots,i_d)\|_F, \\
&\max_{(i_1,\cdots,i_{k+1},i_{k+3},\cdots,i_d)} \|\mathcal{T}(i_1,\cdots,i_{k+1},:,:,i_{k+3},\cdots,i_d)\|_F\} \\
&= \max\left\{ \max_{i_1,\cdots,i_k,i_{k+3},\cdots i_d} \|\mathring{e}_2^{(i_2)} \Diamond \mathcal{T}^0\|_F, \max_{i_1,\cdots,i_k,i_{k+3},\cdots i_d} \|\mathring{e}_{k+2}^{(k+2)} \Diamond \mathcal{T}^0\|_F \right\} \\
&\leq \sqrt{\frac{\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}{n_1^{(1)} L}}.
\end{aligned}
$$

It follows that with high probability

$$
\|\mathcal{P}_{T^{\perp}}(\mathcal{Y})\|_F \leq \frac{6c_4}{\sqrt{c_3}} + \frac{2c_4}{\sqrt{c_3}} \leq \frac{1}{4},
$$

when c_3 large enough.

C. Proof of Lemmas

In this part, we present the proofs of some lemmas mentioned previously.

Proof. Due to

$$
\begin{aligned}
&\langle \mathcal{U} \Diamond \mathcal{U}^{\mathtt{H}} \Diamond e_{i_1\cdots i_d}, e_{i_1\cdots i_d} \rangle \\
=&\langle \mathcal{U} \Diamond \mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(i_1)} \Diamond \dot{h} \Diamond \mathring{e}_2^{(i_2)}, \mathring{e}_1^{(i_1)} \Diamond \dot{h} \Diamond \mathring{e}_2^{(i_2)} \rangle \\
=&\langle \mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(i_1)}, \mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(i_1)} \Diamond \left(\dot{h} \Diamond \mathring{e}_2^{(i_2)} \Diamond \left(\mathring{e}_2^{(i_2)}\right)^{\mathtt{H}} \Diamond \dot{h}^{\mathtt{H}} \right) \rangle \\
=&\langle \mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(i_1)}, \mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(i_1)} \rangle \\
=&\|\mathcal{U} \Diamond \mathring{e}_1^{(i_1)}\|_F^2,
\end{aligned}
$$

where

$$
\dot{h} = \left(\dot{e}_3^{(i_3)}\right) \Diamond \left(\dot{e}_4^{(i_4)}\right) \Diamond \left(\dot{e}_5^{(i_5)}\right) \Diamond \cdots \Diamond \left(\dot{e}_d^{(i_d)}\right).
$$

Since $\mathcal{P}_{\mathbb{T}}$ is self-adjoint, we have

$$
\begin{aligned}
&\|\mathcal{P}_{\mathbb{T}}(e_{i_1\cdots i_d})\|_F^2 \\
=&\langle \mathcal{U} \Diamond \mathcal{U}^{\mathtt{H}} \Diamond e_{i_1\cdots i_d} + e_{i_1\cdots i_d} \Diamond \mathcal{V} \Diamond \mathcal{V}^{\mathtt{H}}, , e_{i_1\cdots i_d} \rangle \\
-&\langle \mathcal{U} \Diamond \mathcal{U}^{\mathtt{H}} \Diamond e_{i_1\cdots i_d} \Diamond \mathcal{V} \Diamond \mathcal{V}^{\mathtt{H}}, e_{i_1\cdots i_d} \rangle \\
=&\|\mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(1)}\|_F^2 + \|\mathcal{V}^{\mathtt{H}} \Diamond \mathring{e}_2^{(i_2)}\|_F^2 - \|\mathcal{U}^{\mathtt{H}} \Diamond \mathring{e}_1^{(1)} \Diamond \dot{h} \Diamond \mathring{e}_2^{(i_2)} \Diamond \mathcal{V}^{\mathtt{H}}\|_F^2 \\
\leq&\frac{\mu \sum_{i=1}^{L} \boldsymbol{R}(i) (n_1 + n_2)}{n_1^{(1)} n_1^{(2)} L^2} \\
\leq&\frac{2\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}{n_1^{(1)} (L)^2}.
\end{aligned}
$$

$\square$

1) The proof of Lemma 7. :

Proof. Assume that $\mathbf{0} \notin \mathcal{Q}$ and consider the problem:

$$\min_{\mathcal{K} \in \mathcal{Q}} \|\mathcal{K}\|_F^2.$$

We may seek approximation from the set

$$\left\{ \mathcal{K} \in \mathcal{Q} : \|\mathcal{K}\|_F^2 < \left\| \mathcal{K}^{\ddagger} \right\|_F^2 \right\},$$

where $\mathcal{K}^{\ddagger} \in \mathcal{Q}$ is arbitrary. Since this set is compact, the existence of a tensor $\mathcal{G} \in \mathcal{Q}$ at which the minimum is obtained is guaranteed. Now if $\mathcal{K} \in \mathcal{Q}$ is arbitrary, by the convexity of $\mathcal{Q}$, then we have

$$\lambda\mathcal{K} + (1-\lambda)\mathcal{G} \in \mathcal{Q}, \quad 0 \le \lambda \le 1.$$

Thus,

$$\begin{aligned} 0 &\le \|\lambda\mathcal{K} + (1-\lambda)\mathcal{G}\|_F^2 - \|\mathcal{G}\|_F^2 \\ &= \|\lambda \operatorname{Hbdiag}(\mathcal{K}_\Phi) + (1-\lambda)\operatorname{Hbdiag}(\mathcal{G}_\Phi)\|_F^2 - \|\operatorname{Hbdiag}(\mathcal{G}_\Phi)\|_F^2 \\ &= \left(\lambda^2 \|\operatorname{Hbdiag}(\mathcal{K}_\Phi) - \operatorname{bdiag}(\mathcal{G}_\Phi)\|_F^2 + 2\lambda \left\langle \operatorname{Hbdiag}(\mathcal{K}_\Phi) - \operatorname{bdiag}(\mathcal{G}_\Phi), \operatorname{Hbdiag}(\mathcal{G}_\Phi)\right\rangle\right) \\ &= 2\lambda\langle \mathcal{K} - \mathcal{G}, \mathcal{G}\rangle + \lambda^2 \|\operatorname{bdiag}(\mathcal{K}_\Phi) - \operatorname{Hbdiag}(\mathcal{G}_\Phi)\|_F^2, \end{aligned}$$

This inequality cannot be valid for small positive λ unless

$$\langle \mathcal{K} - \mathcal{G}, \mathcal{G}\rangle \ge 0, \tag{158}$$

It implies that (158) is equivalent to

$$\langle \mathcal{K}, \mathcal{G}\rangle \ge \langle \mathcal{G}, \mathcal{G}\rangle = \langle \operatorname{Hbdiag}(\mathcal{G}_\Phi), \operatorname{Hbdiag}(\mathcal{G}_\Phi)\rangle = \|\operatorname{Hbdiag}(\mathcal{G}_\Phi)\|_F^2 \ge 0.$$

Since $\mathcal{K}$ is arbitrary, the existence of $\mathcal{K}$ makes the conclusion hold. Now assume that there exists $\mathcal{G} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ such that

$$\langle \mathcal{G}, \mathcal{K}\rangle > 0, \forall \mathcal{K} \in \mathcal{Q}.$$

If $\mathcal{Q}$ contains the origin, this is clearly impossible. □

2) The proof of Lemma 19. :

Proof. The Lemma 19 is proved from the following two aspects.

a) If $\mathcal{X} = \mathbf{0}$, then the result is obvious. Now we assume $\mathcal{X} \neq \mathbf{0}$ and $\mathcal{G} = \mathcal{U}\lozenge\mathcal{V}^{\mathtt{H}} + \mathcal{W}$, where $\mathcal{U}^{\mathtt{H}}\lozenge\mathcal{W} = \mathbf{0}, \mathcal{W}\lozenge\mathcal{V} = \mathbf{0}$ and $\|\mathcal{W}\| \le 1$. It follows that

$$\begin{aligned} \langle \mathcal{G}, \mathcal{X}\rangle &= \left\langle \mathcal{U}\lozenge\mathcal{V}^{\mathtt{H}} + \mathcal{W}, \mathcal{U}\lozenge\mathcal{S}\lozenge\mathcal{V}^{\mathtt{H}}\right\rangle \\ &= \left\langle \mathcal{U}\lozenge\mathcal{V}^{\mathtt{H}}, \mathcal{U}\lozenge\mathcal{S}\lozenge\mathcal{V}^{\mathtt{H}}\right\rangle + \left\langle \mathcal{W}, \mathcal{U}\lozenge\mathcal{S}\lozenge\mathcal{V}^{\mathtt{H}}\right\rangle \\ &= \langle \operatorname{Hbdiag}(\mathcal{I}_\Phi), \operatorname{Hbdiag}(\mathcal{S}_\Phi)\rangle \\ &= \|\operatorname{Hbdiag}((\mathcal{X})_\Phi)\|_* \\ &= \|\mathcal{X}\|_{\mathrm{ETNN}^+}. \end{aligned} \tag{159}$$

Moreover,

$$\|\mathcal{G}\| = \left\|\mathcal{U}\lozenge\mathcal{V}^{\mathtt{H}} + \mathcal{W}\right\| \le \left\|\mathcal{U}\lozenge\mathcal{V}^{\mathtt{H}}\right\| + \|\mathcal{W}\| = 1. \tag{160}$$

Then, equation (160) demonstrates that $\mathcal{G}$ is a subgradient of $\|\cdot\|_{\mathrm{ETNN}^+}$ at $\mathcal{X}$, i.e., $\mathcal{G} \in \partial\|\mathcal{X}\|_{\mathrm{ETNN}^+}$ (see Definition 10).

b) Let $\mathcal{F} := \{\mathcal{U}\lozenge\mathcal{V}^{\mathtt{H}} + \mathcal{W} \mid \mathcal{U}^{\mathtt{H}}\lozenge\mathcal{W} = 0, \mathcal{W}\lozenge\mathcal{V} = 0, \|\mathcal{W}\| \le 1\}$. Suppose that $\mathcal{G} \in \partial\|\mathcal{X}\|_{\mathrm{ETNN}^+}$ but $\mathcal{G} \notin \mathcal{F}$. Then by Lemma 7 there exists $\mathcal{R} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ such that

$$\langle \mathcal{R}, \mathcal{G} - \mathcal{H}\rangle > 0 \text{ for all } \mathcal{H} \in \mathcal{F}, \mathcal{H} - \mathcal{G} \notin \mathbf{0},$$

so that

$$\max_{\mathcal{H} \in \mathcal{F}} \langle \mathcal{R}, \mathcal{H}\rangle < \max_{\mathcal{G} \in \partial\|\mathcal{X}\|_{\mathrm{ETNN}^+}} \langle \mathcal{R}, \mathcal{G}\rangle. \tag{161}$$

The equation (161) is also equivalent to

$$\max_{\mathcal{H} \in \mathcal{F}} \langle \operatorname{Hbdiag}(\mathcal{R}_\Phi), \operatorname{Hbdiag}(\mathcal{H}_\Phi)\rangle < \max_{\mathcal{G} \in \partial\|\mathcal{X}\|_{\mathrm{ETNN}^+}} \langle \operatorname{Hbdiag}(\mathcal{R}_\Phi), \operatorname{Hbdiag}((\mathcal{G})_\Phi)\rangle.$$

Further, let $m = n_1^{(2)} M_1$. For any singular value decomposition, we have

$$\max_{\boldsymbol{D} \in \partial \|\text{Hbdiag}(\mathcal{X}_{k,\Phi})\|_*} \sum_{i=1}^{m} \boldsymbol{D}_i \boldsymbol{U}_i^\top \text{Hbdiag}\left(\mathcal{R}_\Phi\right) \boldsymbol{V}_i < \max_{\text{Hbdiag}(\mathcal{G}_{k,\Phi}) \in \partial \|\text{Hbdiag}(\mathcal{X}_{k,\Phi})\|_*} \langle \text{Hbdiag}\left(\mathcal{R}_\Phi\right), \text{Hbdiag}\left(\mathcal{G}_\Phi\right) \rangle . \quad (162)$$

But the right-hand side or the left-hand side of (162) is just the standard expression for the directional derivative of the convex function $\|\text{Hbdiag}\left(\mathcal{X}_{k,\Phi}\right)\|_*$ in the direction $\text{Hbdiag}\left(\mathcal{R}_\Phi\right)$. Therefore, the equation (162) is contradictory to Lemma 6. The proof of $\mathcal{G} \in \mathcal{F}$ is completed.

□

3) The proof of Lemma 20. :

Proof. For any $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$, $\left(\rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\right) \mathcal{Z}$ can be rewrite

$$\begin{aligned}
\left(\rho^{-1} \mathcal{P}_{\mathbb{T}} \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\right) \mathcal{Z} &= \sum_{i_1 \cdots i_d} \left(\rho^{-1} \delta_{i_1 \cdots i_d} - 1\right) \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}} \mathcal{Z} \rangle \mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right) \\
&:= \sum_{i_1 \cdots i_d} \mathcal{H}_{i_1 \cdots i_d}\left(\mathcal{Z}\right) = \sum_{i_1 \cdots i_d} \bar{\boldsymbol{H}}_{i_1 \cdots i_d}\left(\Phi\left(\mathcal{Z}\right)\right) \\
&= \sum_{i_1 \cdots i_d} \left(\rho^{-1} \delta_{i_1 \cdots i_d} - 1\right) \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}} \mathcal{Z} \rangle \, \text{Hbdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\right)\right),
\end{aligned}$$

where $\mathcal{H}_{i_1 \cdots i_d} : \mathbb{R}^{n \times n \times n_3 \times \cdots \times n_d} \to \mathbb{R}^{\times n \times n \times n_3 \times \cdots \times n_d}$ is a self-adjoint random operator with $\mathbb{E}\left[\mathcal{H}_{i_1 \cdots i_d}\right] = 0$. The matrix operator $\bar{\boldsymbol{H}}_{i_1 \cdots i_d} : \mathbb{B} \to \mathbb{B}$, where $\mathbb{B} = \left\{ \bar{\boldsymbol{B}} : \mathcal{B} \in \mathbb{R}^{n \times n \times n_3 \times \cdots \times n_d} \right\}$ denotes the set consists of block diagonal matrices with the blocks as the frontal slices of $\Phi(\mathcal{B})$. Before using the Bernstein Inequality (Lemma 9) to prove the result of Lemma 20, we need to prove the boundness of $\|\mathcal{H}_{i_1 \cdots i_d}\|$ and $\|\sum_{i_1 \cdots i_d} \mathbb{E}\left[\mathcal{H}_{i_1 \cdots i_d}^2\right]\|$. Based on the above-mentioned definitions, we have $\|\mathcal{H}_{i_1 \cdots i_d}\| = \|\bar{\boldsymbol{H}}_{i_1 \cdots i_d}\|$ and $\|\sum_{i_1 \cdots i_d} \mathbb{E}\left[\mathcal{H}_{i_1 \cdots i_d}^2\right]\| = \|\sum_{i_1 \cdots i_d} \mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1 \cdots i_d}^2\right]\|$. Also $\bar{\boldsymbol{H}}_{i_1 \cdots i_d}$ is self-adjoint and $\mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1 \cdots i_d}\right] = 0$. We only need prove the boundness of $\|\bar{\boldsymbol{H}}_{i_1 \cdots i_d}\|$ and $\|\sum_{i_1 \cdots i_d} \mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1 \cdots i_d}^2\right]\|$.

First,

$$\begin{aligned}
\|\bar{\boldsymbol{H}}_{i_1 \cdots i_d}\| &= \sup_{\|\text{Hbdiag}(\Phi(\mathcal{Z}))\|_F = 1} \|\bar{\boldsymbol{H}}_{i_1 \cdots i_d}\left(\text{Hdiag}\left(\Phi\left(\mathcal{Z}\right)\right)\right)\|_F \\
&\leq \sup_{\|\text{Hdiag}(\Phi(\mathcal{Z}))\|_F = 1} \rho^{-1} \|\mathcal{Z}\|_F \|\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\|_F \|\text{Hdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\right)\right)\|_F \\
&= \sup_{\|\text{Hdiag}(\Phi(\mathcal{Z}))\|_F = 1} \rho^{-1} \|\text{Hdiag}\left(\Phi\left(\mathcal{Z}\right)\right)\|_F \|\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\|_F^2 \\
&\leq \frac{2\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}{n_1^{(1)} (L)^2},
\end{aligned}$$

where the last inequality uses (137).

Second, $\bar{\boldsymbol{H}}_{i_1 \cdots i_d}^2\left(\text{Hbdiag}\left(\Phi\left(\mathcal{Z}_{\langle k \rangle}\right)\right)\right) = \left(\rho^{-1} \delta_{i_1 \cdots i_d} - 1\right)^2 \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}}\left(\mathcal{Z}\right) \rangle \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right) \rangle \text{Hbdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\right)\right)$. Note that $\mathbb{E}\left[\left(\rho^{-1} \delta_{i_1 \cdots i_d} - 1\right)^2\right] \leq \rho^{-1}$, we have

$$\begin{aligned}
&\|\sum_{i_1 \cdots i_d} \mathbb{E}\left[\bar{\boldsymbol{H}}_{i_1 \cdots i_d}^2\left(\text{Hbdiag}\left(\Phi\left(\mathcal{Z}\right)\right)\right)\right]\|_F \\
\leq& \rho^{-1} \|\sum_{i_1 \cdots i_d} \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}}\left(\mathcal{Z}\right) \rangle \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right) \text{Hbdiag}\left(\Phi\left(\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\right)\right) \rangle \|_F \\
\leq& \rho^{-1} \|\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\|_F^2 \|\sum_{i_1 \cdots i_d} \langle e_{i_1 \cdots i_d}, \mathcal{P}_{\mathbb{T}}\left(\mathcal{Z}_{\langle k \rangle}\right) \rangle \|_F \\
=& \rho^{-1} \|\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\|_F^2 \|\mathcal{P}_{\mathbb{T}}\left(\mathcal{Z}\right)\|_F \\
\leq& \rho^{-1} \|\mathcal{P}_{\mathbb{T}}\left(e_{i_1 \cdots i_d}\right)\|_F^2 \|\mathcal{Z}\|_F \\
\leq& \frac{2\mu \sum_{i=1}^{L} \boldsymbol{R}(i)}{n_1^{(1)} L \rho} \|\mathcal{Z}\|_F.
\end{aligned}$$

This implies $\|\sum_{i_1\cdots i_d}\mathbb{E}\left[\bar{\boldsymbol{H}}^2_{i_1\cdots i_d}\left(\text{Hbdiag}\left(\Phi\left(\mathcal{Z}\right)\right)\right)\right]\| \leq \frac{2\mu\sum_{i=1}^{L}\boldsymbol{R}(i)}{n_1^{(1)}L\rho}$. Let $0 \leq \epsilon \leq 1$, based on the Bernstein Inequality (Lemma 9), we have

$$\begin{aligned}&\mathbb{P}\left\{\|\rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}-\mathcal{P}_{\mathbb{T}}\|>\epsilon\right\}\\ =&\mathbb{P}\left\{\|\sum_{i_1\cdots i_d}\mathcal{H}_{i_1\cdots i_d}\|>\epsilon\right\}\\ =&\mathbb{P}\left\{\|\sum_{i_1\cdots i_d}\bar{\boldsymbol{H}}_{i_1\cdots i_d}\|>\epsilon\right\}\\ \leq&2n_1^{(1)}\rho\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2 n_1^{(2)}L\rho}{2\mu\sum_{i=1}^{L}\boldsymbol{R}(i)}\right)\\ =&2\left(n_1^{(1)}L\right)^{1-\frac{3}{16}c_5},\end{aligned}$$

where the last inequality uses $\rho \geq c_4\epsilon^{-2}\frac{\mu\sum_{i=1}^{L}\boldsymbol{R}(i)\log n_1^{(1)}L}{n_1^{(2)}L}$. Thus,

$$\left\|\rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}-\mathcal{P}_{\mathbb{T}}\right\|\leq\epsilon$$

holds with high probability for some numerical constant c_4. □

4) The proof of Lemma 21. :

Proof. For any $\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\cdots,\times n_d}$, $\left(\rho^{-1}\mathcal{P}_{\Omega}-\mathcal{I}\right)\mathcal{Z}$ can be rewrited

$$\begin{aligned}\left(\rho^{-1}\mathcal{P}_{\Omega}-\mathcal{I}\right)\mathcal{Z}&=\sum_{i_1\cdots i_d}\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)\langle e_{i_1\cdots i_d},\mathcal{Z}\rangle e_{i_1\cdots i_d}\\ =\sum_{i_1\cdots i_d}&\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)\mathcal{Z}_{i_1\cdots i_d}\left(\mathring{e}_1^{(i_1)}\right)\lozenge\left(\dot{e}_3^{(i_3)}\right)\left(\dot{e}_4^{(i_4)}\right)\lozenge\cdots\lozenge\left(\dot{e}_d^{(i_d)}\right)\lozenge\left(\mathring{e}_2^{(i_2)}\right)\\ :=\sum_{i_1\cdots i_d}&\mathcal{G}_{i_1\cdots i_d}.\end{aligned}$$

Since $\delta_{i_1\cdots i_d}$ is independent, we have $\mathbb{E}\left[\mathcal{G}_{,i_1\cdots i_d}\right]=0$ and $\|\mathcal{G}_{i_1\cdots i_d}\|\leq\rho^{-1}\|\mathcal{Z}\|$. Moreover,

$$\begin{aligned}&\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{G}_{i_1\cdots i_d}\right)^{\top}\lozenge\mathcal{G}_{i_1\cdots i_d}\right]\|\\ =&\|\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\mathring{e}_2^{(i_2)}\lozenge\left(\mathring{e}_2^{(i_2)}\right)^{\top}\mathbb{E}\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)^2\|\\ =&\|\frac{1-\rho}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\mathring{e}_2^{(i_2)}\lozenge\left(\mathring{e}_2^{(i_2)}\right)^{\top}\|.\end{aligned}$$

Based on the definition of $\mathring{e}_2^{(i_2)}$, the result of $\mathring{e}_2^{(i_2)}\lozenge\left(\mathring{e}_2^{(i_2)}\right)^{\top}$ is an dth-order tensor with its $(1,\cdots,1,i_k,i_{k+1},1,\cdots,1)$-th entry equaling to 1 and the rest equaling to 0. Thus, we can get

$$\begin{aligned}&\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{G}_{i_1\cdots i_d}\right)^{\top}\lozenge\mathcal{G}_{i_1\cdots i_d}\right]\|\\ =&\frac{1-\rho}{\rho}\max_{i_k}\|\sum_{i_1\cdots i_{k-1}i_{k+1}\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\mathring{e}_1^{(i_1)}\left(\mathring{e}_1^{(i_1)}\right)^{\top}\|\\ \leq&\frac{1}{\rho}\|\mathcal{Z}\|^2_{\infty,2},\end{aligned}$$

and $\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{G}_{i_1\cdots i_d}\right)^\top\Diamond\mathcal{G}_{i_1\cdots i_d}\right]\|$ is bounded. Then, using the Bernstein Inequality (Lemma 9), for any $0\leq c\leq 1$, we have

$$
\begin{aligned}
&\|\left(\rho^{-1}\mathcal{P}_\Omega-\mathcal{I}\right)\mathcal{Z}\|=\|\sum_{i_1\cdots i_d}\mathcal{G}_{i_1\cdots i_d}\|\\
\leq&\sqrt{\frac{4c'}{\rho}\|\mathcal{Z}\|_{\infty,2}^2\log\left(2n_1^{(1)}L\right)}+\frac{c'}{\rho}\|\mathcal{Z}\|_\infty\log\left(2n_1^{(1)}L\right)\\
\leq&c_4\left(\frac{\log(n_1^{(1)}L)}{\rho}\|\mathcal{Z}\|_\infty+\sqrt{\frac{\log(n_1^{(1)}L)}{\rho}}\|\mathcal{Z}\|_{\infty,2}\right),
\end{aligned}
$$

holds with high probability for any $c_4\geq\max\{c',2\sqrt{c'}\}$. □

5) The proof of Lemma 12. :

Proof. For any $\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\cdots,\times n_d}$, we define b-th tensor column of $\rho^{-1}\mathcal{P}_\mathbb{T}\mathcal{P}_\Omega\mathcal{Z}-\mathcal{P}_\mathbb{T}\mathcal{Z}$ to be

$$
\begin{aligned}
&\left(\rho^{-1}\mathcal{P}_\mathbb{T}\mathcal{P}_\Omega\mathcal{Z}-\mathcal{P}_\mathbb{T}\mathcal{Z}\right)\Diamond\mathring{e}^{(i_b)}\\
=&\sum_{i_1\cdots i_d}\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)\mathcal{Z}_{i_1\cdots i_d}\mathcal{P}_\mathbb{T}\left(e_{i_1\cdots i_d}\right)\Diamond\mathring{e}_1^{(i_1)}\\
:=&\sum_{i_1\cdots i_d}\mathcal{K}_{i_1\cdots i_d},
\end{aligned}
$$

where $\mathcal{K}_{i_1\cdots i_d}\in\mathbb{R}^{n_1\times 1\times n_3\times\cdots\times n_d}$ are zero-mean independent tensor columns. Based on Lemma 18, we have

$$
\|\mathcal{K}_{i_1\cdots i_d}\|_F=\|\left(\rho^{-1}\delta_{i_1\cdots i_d}-1\right)\mathcal{Z}_{i_1\cdots i_d}\mathcal{P}_\mathbb{T}\left(e_{i_1\cdots i_d}\right)\Diamond\mathring{e}_1^{(i_1)}\|_F\leq\rho^{-1}\sqrt{\frac{2\mu\sum_{i=1}^L\boldsymbol{R}(i)}{n_1^{(1)}L}}\|\mathcal{Z}\|_\infty.
$$

Moreover,

$$
\begin{aligned}
&\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{K}_{i_1\cdots i_d}\right)^\top\Diamond\mathcal{K}_{i_1\cdots i_d}\right]\|_F\\
=&\|\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\left(\mathcal{P}_\mathbb{T}\left(e_{i_1\cdots i_d}\right)\Diamond\mathring{e}_1^{(i_1)}\right)^\top\Diamond\left(\mathcal{P}_\mathbb{T}\left(e_{i_1\cdots i_d}\right)\Diamond\mathring{e}_1^{(i_1)}\right)\mathbb{E}\left(\rho^{-1}\delta_{i_1\cdots i_d}\right)^2\|_F\\
=&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\|\mathcal{P}_\mathbb{T}\left(e_{i_1\cdots i_d}\right)\Diamond\mathring{e}_1^{(i_1)}\|_F^2\\
=&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\|\mathcal{U}\Diamond\mathcal{U}^\top\Diamond e_{i_1\cdots i_d}\Diamond\mathring{e}_1^{(i_1)}+\left(\mathcal{I}-\mathcal{U}\Diamond\mathcal{U}^\top\right)\Diamond e_{i_1\cdots i_d}\Diamond\mathcal{V}\Diamond\mathcal{V}^\top\Diamond\mathring{e}_1^{(i_1)}\|_F^2\\
\leq&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\left(\frac{\sum_{i=1}^L\boldsymbol{R}(i)}{n_1^{(1)}L\mu^{-1}}\|\mathring{e}_2^{(i_2)}\Diamond\mathring{e}_1^{(i_1)}\|_F^2+\|\left(\mathcal{I}-\mathcal{U}\Diamond\mathcal{U}^\top\Diamond\mathring{e}_1^{(i_1)}\Diamond\mathring{e}_1^{(i_1)}\right)\|_F^2\|\mathring{e}_2^{(i_2)}\Diamond\mathcal{V}\Diamond\mathcal{V}^\top\Diamond\mathring{e}_1^{(i_1)}\|_F^2\right)\\
\leq&\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\left(\frac{\mu\sum_{i=1}^L\boldsymbol{R}(i)}{n_1^{(1)}L}\|\mathring{e}_2^{(i_2)}\Diamond\mathring{e}_1^{(i_1)}\|_F^2+\|\mathring{e}_2^{(i_2)}\Diamond\mathcal{V}\Diamond\mathcal{V}^\top\Diamond\mathring{e}_1^{(i_1)}\|_F^2\right)\\
\leq&\frac{\mu\sum_{i=1}^L\boldsymbol{R}(i)}{n_1^{(1)}L\rho}\|\mathcal{Z}\|_{\infty,2}^2+\frac{1}{\rho}\sum_{i_1\cdots i_d}|\mathcal{Z}_{i_1\cdots i_d}|^2\|\mathring{e}_2^{(i_2)}\Diamond\mathcal{V}\Diamond\mathcal{V}^\top\Diamond\mathring{e}_1^{(i_1)}\|_F^2\\
\leq&\frac{2\mu\sum_{i=1}^L\boldsymbol{R}(i)}{n_1^{(1)}L\rho}\|\mathcal{Z}\|_{\infty,2}^2.
\end{aligned}
$$

So, $\|\mathbb{E}\left[\sum_{i_1\cdots i_d}\left(\mathcal{K}_{i_1\cdots i_d}\right)^\top\Diamond\mathcal{K}_{i_1\cdots i_d}\right]\|_F$ is bounded. Based on Bernstein Inequality (Lemma 9), we have

$$
\|\rho^{-1}\mathcal{P}_\mathbb{T}\mathcal{P}_\Omega\mathcal{Z}-\mathcal{P}_\mathbb{T}\mathcal{Z}\|_{\infty,2}\leq\frac{1}{2}\sqrt{\frac{n_1^{(1)}L}{\mu\sum_{i=1}^L\boldsymbol{R}(i)}}\|\mathcal{Z}\|_\infty+\frac{1}{2}\|\mathcal{Z}\|_{\infty,2},
$$

holds with high probability for $m\geq c_3\mu\epsilon^{-2}\sum_{i=1}^L\boldsymbol{R}(i)\,n_1^{(1)}\log\left(n_1^{(1)}L\right)$. □

6) The proof of Lemma 23. :

Proof. To prove this Lemma, we can start from the following two aspects. On the one hand, we deduce

$$\begin{aligned}
&\left\|\rho^{-1}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F^2 \\
&= \left\langle \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{Z}), \mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\rangle \\
&= \left\langle \left(\rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\right)(\mathcal{Z}), \mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\rangle + \left\langle \mathcal{P}_{\mathbb{T}}(\mathcal{Z}), \mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\rangle \\
&\geq \left\|\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F^2 - \left\|\rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\right\| \left\|\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F^2 \\
&\geq \frac{1}{2}\left\|\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F^2,
\end{aligned}$$

where the last inequality utilizes the given condition: $\left\|\rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\right\| \leq \frac{1}{2}$. On the other hand, the condition $\mathcal{P}_{\Omega}(\mathcal{Z}) = 0$ deduces that $\mathcal{P}_{\Omega}(\mathcal{Z}) = 0$, and thus

$$\frac{1}{\sqrt{2}}\left\|\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F \leq \frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F = \frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}^{\perp}}(\mathcal{Z})\right\|_F \leq \frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\mathbb{T}^{\perp}}(\mathcal{Z})\right\|_F \leq \frac{1}{\sqrt{\rho}}\left\|\mathcal{P}_{\mathbb{T}^{\perp}}(\mathcal{Z})\right\|_{\mathrm{ETNN}^+},$$

where the last inequality utilizes

$$\|\mathcal{Z}\|_F = \|\mathrm{Hbdiag}(\mathcal{Z})\|_F \leq \|\mathrm{Hbdiag}(\mathcal{Z})\|_* = \|\mathcal{Z}\|_*.$$

Thus,

$$\left\|\mathcal{P}_{\mathbb{T}}(\mathcal{Z})\right\|_F \leq \sqrt{\frac{2}{\rho}}\left\|\mathcal{P}_{\mathbb{T}^{\perp}}(\mathcal{Z})\right\|_{\mathrm{ETNN}^+}.$$

□

References

[1] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[2] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor completion for estimating missing values in visual data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2012.

[3] Z. Zhang and S. Aeron, “Exact tensor completion using t-SVD,” *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2016.

[4] G. Song, M. K. Ng, and X. Zhang, “Robust tensor completion using transformed tensor singular value decomposition,” *Numerical Linear Algebra with Applications*, vol. 27, no. 3, p. e2299, 2020.

[5] W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo, and T. Huang, “Low-rank high-order tensor completion with applications in visual data,” *IEEE Transactions on Image Processing*, vol. 31, pp. 2433–2448, 2022.

[6] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-Y. Ji, and T.-H. Ma, “Tensor N-tubal rank and its convex relaxation for low-rank tensor recovery,” *Information Sciences*, vol. 532, pp. 170–189, 2020.

[7] E. Kernfeld, M. Kilmer, and S. Aeron, “Tensor–tensor products with invertible linear transforms,” *Linear Algebra and its Applications*, vol. 485, pp. 545–570, 2015.

[8] K. Fan and A. J. Hoffman, “Some metric inequalities in the space of matrices,” *Proceedings of the American Mathematical Society*, vol. 6, no. 1, pp. 111–116, 1955.

[9] G. A. Watson, “Characterization of the subdifferential of some matrix norms,” *Linear Algebra Appl*, vol. 170, no. 1, pp. 33–45, 1992.

[10] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1997, vol. 11.

[11] J. A. Tropp, “User-friendly tail bounds for sums of random matrices,” *Foundations of computational mathematics*, vol. 12, pp. 389–434, 2012.

[12] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018, vol. 47.

[13] X. Li, “Compressed sensing and matrix completion with constant proportion of corruptions,” *Constructive Approximation*, vol. 37, pp. 73–99, 2013.

[14] E. Candes and B. Recht, “Exact matrix completion via convex optimization,” *Communications of the ACM*, vol. 55, no. 6, pp. 111–119, 2012.

[15] D. Gross, “Recovering low-rank matrices from few coefficients in any basis,” *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[16] B. Recht, “A simpler approach to matrix completion.” *Journal of Machine Learning Research*, vol. 12, no. 12, 2011.

[17] Y. Chen, “Incoherence-optimal matrix completion,” *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2909–2923, 2015.

[18] Y. Wang, W. Yin, and J. Zeng, “Global convergence of admm in nonconvex nonsmooth optimization,” *Journal of Scientific Computing*, vol. 78, pp. 29–63, 2019.

[19] 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.

[20] 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.