

# SUPPLEMENTARY MATERIAL OF HIGH-ORDER TENSOR RECOVERY WITH A TENSOR $U_1$ NORM

**Anonymous authors**

Paper under double-blind review

## 1 RELATED WORKS

Unlike the well-defined matrix rank, there is currently no universally accepted definition for tensor rank. In this part, we will briefly introduce three common definitions of tensor rank based on different decomposition techniques: Canonical Polyadic (CP) Decomposition Hitchcock (1927; 1928), Tucker Decomposition Tucker (1963), and methods based on t-SVD Lu et al. (2019b; 2018); Qin et al. (2022); Wang et al. (2023); Zhang et al. (2022); Zhou et al. (2017).

It is worth noting that from an equivalent definition of matrix rank, a rank  $r$  matrix can be written as the sum of  $r$  rank-one matrix. Inspired by that, Kolda and Bader Kolda & Bader (2009b) have proposed CP rank, *i.e.*,  $\text{rank}_{\text{cp}}(\cdot)$ , defined on tensor rank-one decomposition (CP Decomposition):

$$\text{rank}_{\text{cp}}(\mathcal{X}) = \min\{R \mid \mathcal{X} = \sum_{r=1}^R g_{r,r,\dots,r} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \dots \circ \mathbf{u}_r^{(h)}, \mathbf{u}_r^{(j)} \in \mathbb{R}^{I_j} \text{ for } j = 1, 2, \dots, h\} \quad (1)$$

for tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_h}$ . We can see from (1) that the definition of matrix rank is a special case of CP rank. But solving (1) is time-consuming even for small tensor when  $h \geq 3$ .

As the computation of CP rank is NP-hard and greatly restricts its application in tensor recovery, various Tucker Decomposition-based methods for defining tensor rank have been proposed and extensively studied than CP rank Kolda & Bader (2009b). Given  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_h}$ , the Tucker Decomposition of  $\mathcal{A}$  is written as  $\mathcal{A} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \dots \times_h \mathbf{U}_h$ , where  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_h}$ , and  $\mathbf{U}_k \in \mathbb{R}^{I_k \times R_k}$  for  $k = 1, 2, \dots, h$ . Given  $\text{rank}(\mathbf{A}_{(k)})$  for all  $k$ , we can obtain the decomposition by the higher-order singular value decomposition (HOSVD) De Lathauwer et al. (2000), where  $\text{rank}(\mathbf{A}_{(k)}) = R_k$  for  $k = 1, 2, \dots, h$ . Therefore, the Tucker rank of tensor  $\mathcal{A}$  is defined as

$$\text{rank}_{\text{tc}}(\mathcal{A}) = (\text{rank}(\mathbf{A}_{(1)}), \text{rank}(\mathbf{A}_{(2)}), \dots, \text{rank}(\mathbf{A}_{(h)}),$$

which is also known as multilinear rank and n-rank. Based on the Tucker rank, Gandy *et al.* given a new rank of the tensor data defined as  $\sum_{n=1}^h \text{rank}(\mathbf{A}_{(n)})$  Gandy et al. (2011). Furthermore, considering the difference of the low rankness along different dimensions of tensor data, Liu et al. (2013) give a weighted sum of the ranks of the unfolding matrices  $\sum_{n=1}^h \alpha_n \text{rank}(\mathbf{A}_{(n)})$ , where the weights  $\alpha_n (n = 1, 2, \dots, h)$  satisfy  $\sum_{n=1}^h \alpha_n = 1$  and play an important role in the newly defined rank. However, the best choice for the weights is hard to know if without any prior. Thus, a new tensor rank based on the maximum rank of a set of unfolding matrices is proposed to promote the low-rankness of unfolding matrices of the recovered tensor Zhang et al. (2014a).

Recently, there has been a growing interest in tensor rank by using t-SVD Zhang et al. (2014b); Lu et al. (2019a;b). This approach defines rank based on the Singular Value Decomposition (SVD) of frontal slices of the tensor resulting from invertible transforms applied along a specific dimension (known as t-SVD). This approach is widely employed in tensor recovery, as they can better utilize the smoothness priors in tensor data due to the use of transforms such as DFT. For example, Zhang et al. (2014b) introduced a tensor tubal rank based on the Discrete Fourier Transform (DFT) for three-order tensors. It counts the number of non-zero tensor tubes in the singular value tensor obtained by performing frontal-slices-wise SVD of the transformed tensor. Similarly, Lu et al. (2019a) defined tensor average rank for three-order tensors based on DFT, which averages the ranks of frontal slices of the transformed tensor and provided theoretical guarantees for exact recovery using the convex hull of tensor average rank. As noted in Lu et al. (2019a), the low tensor average rank assumption for tensor data can be more easily satisfied in the real world than the low-rank assumption employed in the

---

**Algorithm 2:** Tensor Decomposition Based on Slices-Wise Low-Rank Prior (TDSL)

---

**Input:**  $\mathcal{A} \in \mathbb{R}^{I_{k_1} \times I_{k_2} \times \dots \times I_{k_h}}$ ,  $\{\hat{U}_{k_n}\}_{n=3}^s$ , and  $r$ , where  $1 \leq k_i \neq k_j$  (if  $i \neq j$ )  $\leq h$

**Output:**  $\mathcal{Z}_1$ ,  $\{U_{k_n}\}_{n=s+1}^h$

1.  $\bar{\mathcal{A}} = \mathcal{A} \times_{k_3} \hat{U}_{k_3} \cdots \times_{k_s} \hat{U}_{k_s}$

**while** not converged **do**

2. Calculate the slices-wise SVD for  $\bar{\mathcal{A}} \times_{k_{s+1}} U_{k_{s+1}}^{(t)} \cdots \times_{k_h} U_{k_h}^{(t)}$  by computing SVD of its all slices along the  $(k_1, k_2)$ -th mode: for all  $1 \leq i_{k_3} \leq I_{k_3}, \dots, 1 \leq i_{k_h} \leq I_{k_h}$ , we have

$$[\bar{\mathcal{A}} \times_{k_{s+1}} U_{k_{s+1}}^{(t)} \cdots \times_{k_h} U_{k_h}^{(t)}]_{:, :, i_{k_3}, \dots, i_{k_h}} = [\bar{\mathcal{U}}]_{:, :, i_{k_3}, \dots, i_{k_h}} [\bar{\mathcal{S}}]_{:, :, i_{k_3}, \dots, i_{k_h}} [\bar{\mathcal{V}}]_{:, :, i_{k_3}, \dots, i_{k_h}}^T.$$

3. Calculate  $\mathcal{Z}_1^{(t+1)}$  by  $[\mathcal{Z}_1]_{:, :, i_{k_3}, \dots, i_{k_h}}^{(t+1)} = [\bar{\mathcal{U}}]_{:, 1:r, i_{k_3}, \dots, i_{k_h}} [\bar{\mathcal{S}}]_{1:r, 1:r, i_{k_3}, \dots, i_{k_h}} [\bar{\mathcal{V}}]_{:, 1:r, i_{k_3}, \dots, i_{k_h}}^T$

4. Compute  $U_{k_n}^{(t+1)}$  for all  $s+1 \leq n \leq h$  by  $U_{k_n}^{(t+1)} = UV^T$ , where  $U$  and  $V$  are obtained by SVD for  $[\mathcal{Z}_1]_{(k_n)} \mathcal{Y}_{(k_n)}^T$ , i.e.,  $[\mathcal{Z}_1]_{(k_n)} \mathcal{Y}_{(k_n)}^T = USV^T$ , and

$$\mathcal{Y} = \bar{\mathcal{A}} \times_{k_{s+1}} U_{k_{s+1}}^{(t+1)} \cdots \times_{k_{n-1}} U_{k_{n-1}}^{(t+1)} \times_{k_{n+1}} U_{k_{n+1}}^{(t)} \cdots \times_{k_h} U_{k_h}^{(t)}.$$

3. Check the convergence conditions:  $\|\mathcal{Z}_1^{(t+1)} - \mathcal{Z}_1^{(t)}\|_\infty < \varepsilon$ ,  $\|U_{k_n}^{(t+1)} - U_{k_n}^{(t)}\|_\infty < \varepsilon$  for all  $s+1 < n \leq h$ ;

4.  $t = t + 1$ .

**end while**

---

tensor tubal rank, CP rank, and Tucker rank. Specifically, the tensor average rank of any three-order tensor  $\mathcal{A}$  satisfied the following inequation

$$\text{rank}_a(\mathcal{A}) \leq \max \text{rank}_{\text{tc}}(\mathcal{A}) \leq \text{rank}_{\text{cp}}(\mathcal{A}), \quad (2)$$

where  $\text{rank}_{\text{tc}}(\mathcal{A})$  and  $\text{rank}_{\text{cp}}(\mathcal{A})$  are the Tucker rank Kolda & Bader (2009a) and CP rank Kolda & Bader (2009b) of  $\mathcal{A}$ , respectively. Employing a similar idea to tensor average rank, a new rank based on real invertible transforms has been given in Lu et al. (2019b), and defined as  $\text{rank}_{\mathcal{L}}(\mathcal{A}) = \frac{1}{I_3} \sum_{i_3=1}^{I_3} \text{rank}([\mathcal{A} \times_3 \mathcal{L}]_{:, :, i_3})$ , where  $\mathcal{L}$  is a fixed real invertible transform, such as Discrete Cosine Matrix (DCM) and Random Orthogonal Matrix (ROM), that satisfies  $\mathcal{L}^T \mathcal{L} = \mathcal{L} \mathcal{L}^T = \ell_{\mathcal{L}}$ , and  $\ell_{\mathcal{L}}$  is a constant. To handle the higher order tensor case, in Qin et al. (2022), the slice-wise low rankness of  $\mathcal{L}(\mathcal{A})$  is considered, where  $\mathcal{L}(\mathcal{A}) = \mathcal{X} \times_3 \mathcal{L}_3 \times_4 \cdots \times_h \mathcal{L}_h$ ,  $\mathcal{L}^T(\mathcal{A}) = \mathcal{X} \times_h \mathcal{L}_h^T \times_{h-1} \cdots \times_3 \mathcal{L}_3^T$ , and  $\mathcal{L}^T(\mathcal{L}(\mathcal{I})) = \mathcal{L}(\mathcal{L}^T(\mathcal{I})) = \ell_{\mathcal{L}} \mathcal{I}$  for given invertible transforms  $\{\mathcal{L}_k\}_{k=3}^h$ . Considering the difference of tensor low-rankness across different dimensions of the tensor, Zheng et al. (2020) give WSTNN, which is defined as the weighted sum of the tensor average rank of all  $\binom{h}{2}$  mode- $k_1 k_2$  unfolding tensor. However, it will become impractical as the tensor order  $h$  increases. Besides, the weight parameter tuning can also be a challenge. These t-SVD-based methods utilize the smoothness priors in tensor data better than the other methods due to the use of transforms such as DFT, but it is also exactly why they are sensitive to non-smooth changes and slice permutations of tensor data. Zheng et al. (2022) proposed a solution to address the slice permutation issue in DFT-based methods by minimizing a Hamiltonian circle, though it is limited to DFT. Moreover, the methods based on t-SVD introduce more variables and weight parameters compared to CP and Tucker rank methods.

## 2 TDSL (ALGORITHM 2)

## 3 THE PROOF OF PROPERTY 2

*Proof.* (i) We can conclude that both the tensor  $U_1$  norm and tensor  $U_\infty$  norm are convex due to the convexity properties of the  $l_1$ -norm and  $\infty$ -norm, respectively.

(ii)

$$\begin{aligned} \sup_{\|\mathcal{B}\|_{U, \infty} \leq 1} \langle \mathcal{A}, \mathcal{B} \rangle &= \sup_{\|\mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h\|_\infty \leq 1} \langle \mathcal{A}, \mathcal{B} \rangle \\ &= \sup_{\|\mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h\|_\infty \leq 1} \left\langle \mathcal{A} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h, \mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h \right\rangle. \end{aligned}$$

Let  $\hat{\mathcal{B}} = \mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h$  be any tensor. Then we have

$$\sup_{\|\hat{\mathcal{B}}\|_\infty \leq 1} \langle \mathcal{A} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h, \hat{\mathcal{B}} \rangle = \|\mathcal{A} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h\|_1 = \|\mathcal{A}\|_{\mathcal{U},1}. \quad (3)$$

(iii) The proof is completed in the following two steps, utilizing the properties of conjugate functions presented in Fazel Sarjoui (2002); Hiriarturruty & Lemaréchal (1993), *i.e.*, the conjugate of the conjugate,  $\phi_0^{**}$ , is the convex envelope of a given function  $\phi_0 : \mathbb{C} \rightarrow \mathbb{R}$ . For given function  $\phi_0$ , the conjugate  $\phi_0^*$  of the function  $\phi_0$  is defined as  $\phi_0^*(y) = \sup\{\langle y, x \rangle - \phi_0(x) | x \in \mathbb{C}\}$ .

STEP1. COMPUTING THE CONJUGATE OF SPARSITY-BASED TENSOR  $U_0, \phi^*$ .

$$\begin{aligned} \phi^*(\mathcal{B}) &= \sup_{\mathcal{A} \in \mathbb{S}} \langle \mathcal{B}, \mathcal{A} \rangle - \|\mathcal{A}\|_{\mathcal{U},0} = \sup_{\|\mathcal{A}\|_{\mathcal{U},\infty} \leq 1} \langle \mathcal{B}, \mathcal{A} \rangle - \|\mathcal{A} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h\|_0 \\ &\quad (\text{Let } \hat{\mathcal{A}} = \mathcal{A} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h \text{ be any tensor.}) \\ &= \sup_{\|\hat{\mathcal{A}}\|_\infty \leq 1} \langle \mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h, \hat{\mathcal{A}} \rangle - \|\hat{\mathcal{A}}\|_0 \\ &= \begin{cases} 0, & \|\mathcal{B}\|_{\mathcal{U},\infty} \leq 1; \\ \|(\mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h, 1)_+\|_1, & \text{otherwise.} \end{cases} \end{aligned}$$

STEP2. COMPUTING THE CONJUGATE OF  $\phi^*, \phi^{**}$ . Defining

$$f(\mathcal{A}_0) = \begin{cases} 0, & \|\mathcal{A}_0\|_\infty \leq 1; \\ \|(\mathcal{A}_0, 1)_+\|_1, & \text{otherwise,} \end{cases}$$

we have

$$\begin{aligned} \phi^{**}(\mathcal{C}) &= \sup_{\mathcal{B}} \langle \mathcal{C}, \mathcal{B} \rangle - \phi^*(\mathcal{B}) = \sup_{\mathcal{B}} \langle \mathcal{C} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h, \mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h \rangle - \phi^*(\mathcal{B}) \\ &\quad (\text{Let } \hat{\mathcal{B}} = \mathcal{B} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h \text{ be any tensor.}) \\ &= \sup_{\hat{\mathcal{B}}} \langle \mathcal{C} \times_1 \hat{U}_1 \times_2 \cdots \times_h \hat{U}_h, \hat{\mathcal{B}} \rangle - f(\hat{\mathcal{B}}) = \|\mathcal{C}\|_{\mathcal{U},1} \end{aligned}$$

over the set  $\mathbb{S}$ .

□

## 4 THE PROOF OF THEOREM 1

Without loss of generality, let us consider

$$\min_{\mathcal{Z}, U_k^T U_k = I(k=s+1, \dots, h)} \|\mathcal{Z}\|_{\mathcal{U},1} \quad \text{s.t. } \Psi_{\mathbb{I}}(\mathcal{M}) = \mathcal{Z} \times_{s+1} U_{s+1}^T \cdots \times_h U_h^T + \mathcal{E}, \quad (4)$$

where  $\mathcal{U}(\mathcal{Z}) = \mathcal{Z} \times_1 U_1 \cdots \times_s U_s$ .

$$\begin{aligned} &\mathcal{L}_a(\mathcal{Z}, \{U_k\}_{k=s+1}^h, \mathcal{E}, \mathcal{Y}, \{Y_k\}_{k=s+1}^h, \mathcal{W}) \\ &= \|\mathcal{Z}\|_{\mathcal{U},1} + \langle \Psi_{\mathbb{I}}(\mathcal{M}) - \mathcal{Z} \times_h U_h^T \times_2 \cdots \times_{s+1} U_{s+1}^T - \mathcal{E}, \mathcal{Y} \rangle + \sum_{k=s+1}^h \langle U_k^T U_k - I, Y_k \rangle + \langle \Psi_{\mathbb{I}}(\mathcal{E}), \mathcal{W} \rangle \end{aligned} \quad (5)$$

From (5), *i.e.*, the Lagrangian function of (4), we can get the following KKT conditions by the first order optimality conditions for (4):

$$\begin{cases} \Psi_{\mathbb{I}}(\mathcal{M}) - \mathcal{X} - \mathcal{E} = \mathbf{0}; \\ \mathcal{Y} \times_{s+1} U_{s+1} \times_{s+2} \cdots \times_h U_h \in \partial \|\mathcal{Z}\|_{\mathcal{U},1}; \\ U_k^T U_k = I \text{ for } k = s+1, s+2, \dots, h \\ -\mathcal{F}_{(k)}(\mathcal{C}_{(k)})^T + U_k(Y_k + Y_k^T) = \mathbf{0}; \\ \Psi_{\mathbb{I}}(\mathcal{E}) = \mathbf{0}; \\ -\Psi_{\mathbb{I}^c}(\mathcal{Y}) = \mathbf{0}; \\ -\Psi_{\mathbb{I}}(\mathcal{Y}) + \mathcal{W} = \mathbf{0}, \end{cases} \quad (6)$$

where  $\mathcal{C} = \mathcal{Y} \times_{s+1} U_{s+1} \cdots \times_{k-1} U_{k-1}$  and  $\mathcal{F} = \mathcal{Z} \times_h (U_h)^T \cdots \times_{k+1} (U_{k+1})^T$ .

*Proof.* (i) By  $\Psi_{\mathbb{I}}(\mathcal{M}) - \mathcal{X}^{(t+1)} - \boldsymbol{\varepsilon}^{(t+1)} = (\mu^{(t)})^{(-1)}(\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)})$  and the boundedness of  $\mathbf{y}^{(t)}$ , we have  $\lim_{t \rightarrow \infty} \Psi_{\mathbb{I}}(\mathcal{M}) - \mathcal{X}^{(t+1)} - \boldsymbol{\varepsilon}^{(t+1)} = \mathbf{0}$ .

(ii) From the the optimality of  $\mathcal{Z}^{(t+1)}$ ,  $\{\mathbf{U}_k^{(t+1)}\}_{k=s+1}^h$ , and  $\boldsymbol{\varepsilon}^{(t+1)}$ , we have

$$\begin{aligned}
& \mathcal{L}(\mathcal{Z}^{(t+1)}, \{\mathbf{U}_k^{(t+1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t+1)}, \mathbf{y}^{(t)}, \mu^{(t)}) \\
& \leq \mathcal{L}(\mathcal{Z}^{(t+1)}, \{\mathbf{U}_k^{(t+1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t+1)}, \mathbf{y}^{(t)}, \mu^{(t)}) + \frac{\eta^{(t)}}{2} \|\mathcal{Z}^{(t+1)} - \mathcal{Z}^{(t)}\|_F^2 \\
& \quad + \frac{\eta^{(t)}}{2} \sum_{k=s+1}^h \|\mathbf{U}_k^{(t+1)} - \mathbf{U}_k^{(t)}\|_F^2 + \frac{\eta^{(t)}}{2} \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(t)}, \{\mathbf{U}_k^{(t+1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t+1)}, \mathbf{y}^{(t)}, \mu^{(t)}) + \frac{\eta^{(t)}}{2} \sum_{k=s+1}^h \|\mathbf{U}_k^{(t+1)} - \mathbf{U}_k^{(t)}\|_F^2 \\
& \quad + \frac{\eta^{(t)}}{2} \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(t)}, \{\mathbf{U}_k^{(t)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t+1)}, \mathbf{y}^{(t)}, \mu^{(t)}) + \frac{\eta^{(t)}}{2} \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(t)}, \{\mathbf{U}_k^{(t)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t)}, \mathbf{y}^{(t)}, \mu^{(t)}) \\
& = \mathcal{L}(\mathcal{Z}^{(t)}, \{\mathbf{U}_k^{(t)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t)}, \mathbf{y}^{(t-1)}, \mu^{(t-1)}) + \frac{1}{2} (\mu^{(t-1)})^{-2} (\mu^{(t-1)} + \mu^{(t)}) \|\mathbf{y}^{(t)} - \mathbf{y}^{(t-1)}\|_F^2.
\end{aligned} \tag{7}$$

Therefore, we have

$$\begin{aligned}
\|\mathcal{Z}^{(t+1)}\|_{\mathcal{U},1} & \leq \mathcal{L}(\mathcal{Z}^{(t+1)}, \{\mathbf{U}_k^{(t+1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t+1)}, \mathbf{y}^{(t)}, \mu^{(t)}) + \|\mathbf{y}^{(t)}\|_F^2 / (\mu^{(t)})^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(t)}, \{\mathbf{U}_k^{(t)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t)}, \mathbf{y}^{(t)}, \mu^{(t)}) + \|\mathbf{y}^{(t)}\|_F^2 / (\mu^{(t)})^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(1)}, \{\mathbf{U}_k^{(1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(1)}, \mathbf{y}^{(0)}, \mu^{(0)}) \\
& \quad + \frac{1}{2} \sum_{n=1}^t (\mu^{(n-1)})^{-2} (\mu^{(n-1)} + \mu^{(n)}) \|\mathbf{y}^{(n)} - \mathbf{y}^{(n-1)}\|_F^2 + \|\mathbf{y}^{(t)}\|_F^2 / (\mu^{(t)})^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(1)}, \{\mathbf{U}_k^{(1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(1)}, \mathbf{y}^{(0)}, \mu^{(0)}) + \sum_{n=1}^t (\mu^{(n-1)})^{-2} \mu^{(n)} \|\mathbf{y}^{(n)} - \mathbf{y}^{(n-1)}\|_F^2 \\
& \quad + \|\mathbf{y}^{(t)}\|_F^2 / (\mu^{(t)})^2.
\end{aligned} \tag{8}$$

From (8),  $\sum_{t=1}^{\infty} (\mu^{(t)})^{-2} \mu^{(t+1)} < +\infty$ , and the boundedness of  $\mathbf{y}^{(t)}$ , we can know that  $\mathcal{Z}^{(t)}$  is bounded. Besides, since  $\|\mathbf{U}_k^{(t)}\|_F = \sqrt{I_k}$  holds for any positive integer  $t$ ,  $\mathbf{U}_k^{(t)}$  and  $\mathcal{X}^{(t)}$  are bounded. Therefore,  $\boldsymbol{\varepsilon}^{(t)}$  is bounded from  $\lim_{t \rightarrow \infty} \Psi_{\mathbb{I}}(\mathcal{M}) - \mathcal{X}^{(t+1)} - \boldsymbol{\varepsilon}^{(t+1)} = \mathbf{0}$ .

(iii) From (7), we have

$$\begin{aligned}
& \sum_{t=1}^n \frac{\eta^{(t)}}{2} (\|\mathcal{Z}^{(t+1)} - \mathcal{Z}^{(t)}\|_F^2 + \sum_{k=s+1}^h \|\mathbf{U}_k^{(t)} - \mathbf{U}_k^{(t+1)}\|_F^2 + \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2) \\
& \quad - \sum_{t=1}^n \frac{1}{2} (\mu^{(t-1)})^{-2} (\mu^{(t-1)} + \mu^{(t)}) \|\mathbf{y}^{(t)} - \mathbf{y}^{(t-1)}\|_F^2 \\
& \leq \mathcal{L}(\mathcal{Z}^{(1)}, \{\mathbf{U}_k^{(1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(1)}, \mathbf{y}^{(0)}, \mu^{(0)}) - \mathcal{L}(\mathcal{Z}^{(n+1)}, \{\mathbf{U}_k^{(n+1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(n+1)}, \mathbf{y}^{(n)}, \mu^{(n)}) \\
& \leq \mathcal{L}(\mathcal{Z}^{(1)}, \{\mathbf{U}_k^{(1)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(1)}, \mathbf{y}^{(0)}, \mu^{(0)}) + \|\mathbf{y}^{(n)}\|_F^2 / (\mu^{(n)})^2
\end{aligned} \tag{9}$$

Since  $\mathbf{Y}^{(n)}$  is bounded, there exists  $M_0$  and  $M_1$  such that

$$\begin{aligned} & \sum_{t=1}^n \frac{\eta^{(t)}}{2} (\|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F^2 + \sum_{k=s+1}^h \|\mathbf{U}_k^{(t+1)} - \mathbf{U}_k^{(t)}\|_F^2 + \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2) \\ & \leq M_0 + \sum_{t=1}^n \frac{1}{2} (\mu^{(t-1)})^{-2} (\mu^{(t-1)} + \mu^{(t)}) M_1 \leq M_0 + \sum_{t=1}^n (\mu^{(t-1)})^{-2} \mu^{(t)} M_1. \end{aligned} \quad (10)$$

As  $n$  approaches infinity, we have

$$\begin{aligned} & \sum_{t=1}^{\infty} \frac{\eta^{(t)}}{2} (\|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F^2 + \sum_{k=s+1}^h \|\mathbf{U}_k^{(t+1)} - \mathbf{U}_k^{(t)}\|_F^2 + \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2) \\ & \leq M_0 + \sum_{t=1}^{\infty} (\mu^{(t-1)})^{-2} \mu^{(t)} M_1 < \infty. \end{aligned} \quad (11)$$

(iv) From (iii) we can see that there exists  $M_2$  such that

$$\max(\|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F^2, \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F^2, \{\|\mathbf{U}_k^{(t+1)} - \mathbf{U}_k^{(t)}\|_F^2\}_{k=s+1}^h) \leq (\eta^{(t)})^{(-1)} M_2^2,$$

therefore

$$\begin{aligned} & \|\boldsymbol{\chi}^{(t+1)} - \boldsymbol{\chi}^{(t)}\|_F \\ & = \|\mathbf{Z}^{(t+1)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t+1)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t+1)})^T \\ & \quad + \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t+1)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T\|_F \\ & \leq \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F + \|\mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t+1)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T\|_F \\ & = \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F + \|\mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t+1)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T\|_F \\ & \quad + \|\mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T\|_F \\ & \leq \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F + \|\mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t+1)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T\|_F \\ & \quad + \|\mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t)})^T \cdots \times_{s+1} (\mathbf{U}_{s+1}^{(t)})^T\|_F \\ & \leq \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F + \|\mathbf{U}_{s+1}^{(t+1)} - \mathbf{U}_{s+1}^{(t)}\|_F \|\mathbf{Z}^{(t)}\|_F + \|\mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t+1)})^T \cdots \times_{s+2} (\mathbf{U}_{s+2}^{(t+1)})^T \\ & \quad - \mathbf{Z}^{(t)} \times_h (\mathbf{U}_h^{(t)})^T \cdots \times_{s+2} (\mathbf{U}_{s+2}^{(t)})^T\|_F \\ & \leq \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_F + \sum_{k=s+1}^h (\|\mathbf{U}_k^{(t+1)} - \mathbf{U}_k^{(t)}\|_F) \|\mathbf{Z}^{(t)}\|_F \\ & \leq (\eta^{(t)})^{(-1/2)} (1 + h \|\mathbf{Z}^{(t)}\|_F) M_2. \end{aligned} \quad (12)$$

From the boundedness of  $\mathbf{Z}^{(t)}$ , there exists  $M_3$  such that  $\|\boldsymbol{\chi}^{(t+1)} - \boldsymbol{\chi}^{(t)}\|_F^2 \leq (\eta^{(t)})^{(-1)} M_3$ .

Let  $\mathcal{D}^{(t+1)} = \Psi_{\mathbb{I}}(\mathcal{M}) - \boldsymbol{\chi}^{(t+1)} - \boldsymbol{\varepsilon}^{(t+1)}$ . From the above discussion, we know that there exists  $M_4$  such that  $\|\mathcal{D}^{(t+1)} - \mathcal{D}^{(t)}\|_F \leq \|\boldsymbol{\chi}^{(t+1)} - \boldsymbol{\chi}^{(t)}\|_F + \|\boldsymbol{\varepsilon}^{(t+1)} - \boldsymbol{\varepsilon}^{(t)}\|_F \leq (\eta^{(t)})^{(-1/2)} M_4$ . Thus, we have

$$\|\mathcal{D}^{(t)}\|_F \leq (\eta^{(t)})^{(-1/2)} M_4 + \|\mathcal{D}^{(t+1)}\|_F \leq M_4 \sum_{n=0}^m (\eta^{(t+n)})^{(-1/2)} + \|\mathcal{D}^{(t+1+m)}\|_F$$

for any  $m > 0$  and  $(\mu^{(n)})^{(-1)} \|\mathbf{Y}^{(n+1)} - \mathbf{Y}^{(n)}\|_F = \|\mathcal{D}^{(n+1)}\|_F \leq M_4 \sum_{t=n+1}^{+\infty} (\eta^{(t)})^{(-1/2)}$  when  $m \rightarrow \infty$ .

From  $\lim_{n \rightarrow \infty} \mu^{(n)} \sum_{t=n}^{\infty} (\eta^{(t)})^{-1/2} = 0$ , we have  $\lim_{n \rightarrow \infty} \|\mathbf{Y}^{(n+1)} - \mathbf{Y}^{(n)}\|_F = 0$ .

(v) From the boundedness of  $\{\{\mathbf{Z}^{(t)}, \{\mathbf{U}_k^{(t)}\}_{k=s+1}^h, \boldsymbol{\chi}^{(t)}, \boldsymbol{\varepsilon}^{(t)}\}\}$ , there exist a subsequence  $\{\{\mathbf{Z}^{(t_i)}, \{\mathbf{U}_k^{(t_i)}\}_{k=s+1}^h, \boldsymbol{\varepsilon}^{(t_i)}, \boldsymbol{\chi}^{(t_i)}\}\}$  and  $[\mathbf{Z}^*, \{\mathbf{U}_k^*\}_{k=s+1}^h, \boldsymbol{\varepsilon}^*, \boldsymbol{\chi}^*]$  such that

$\lim_{i \rightarrow +\infty} [\mathcal{Z}^{(t_i)}, \{U_k^{(t_i)}\}_{k=s+1}^h, \mathcal{E}^{(t_i)}, \mathcal{Y}^{(t_i)}] = [\mathcal{Z}^*, \{U_k^*\}_{k=s+1}^h, \mathcal{E}^*, \mathcal{Y}^*]$ . From the optimality of  $\mathcal{Z}^{(t_i+1)}$  and the convexity of the tensor  $U_1$  norm, there exists  $\mathcal{H}^{(t_i+1)} \in \partial \|\mathcal{Z}^{(t_i+1)}\|_{\mathcal{U},1}$  such that

$$\mathcal{H}^{(t_i+1)} + \mu^{(t_i)} (\mathcal{Z}^{(t_i+1)} - \mathcal{P}^{(t_i)} \times_{s+1} U_{s+1}^{(t_i)} \times_{s+2} \cdots \times_h U_h^{(t_i)}) + \eta^{(t_i)} (\mathcal{Z}^{(t_i+1)} - \mathcal{Z}^{(t_i)}) = \mathbf{0}$$

and

$$\mathcal{H}^* - \mathcal{Y}^* \times_{s+1} U_{s+1}^* \times_{s+2} \cdots \times_h U_h^* = \mathbf{0},$$

where  $\lim_{i \rightarrow +\infty} \mathcal{H}^{(t_i+1)} = \mathcal{H}^*$ , and  $\hat{\mathcal{P}}^{(t_i)} = \Psi(\mathcal{M}) - \mathcal{E}^{(t_i)} + \frac{1}{\mu^{(t_i)}} \mathcal{Y}^{(t_i)}$ . By the upper semi-continuous property of the subdifferential Clarke (1983),  $\mathcal{Y}^* \times_{s+1} U_{s+1}^* \times_{s+2} \cdots \times_h U_h^* = \mathcal{H}^* \in \partial \|\mathcal{Z}^*\|_{\mathcal{U},1}$ .

From the optimality of  $U_k^{(t_i+1)}$ , we have  $(U_k^{(t_i+1)})^T U_k^{(t_i+1)} = I$ , and there exists  $\mathbf{Y}_k^{(t_i+1)}$  such that  $\mathbf{0} = \mu^{(t_i)} (U_k^{(t_i+1)} \mathcal{B}_{(k)} - \mathcal{A}_{(k)}) \mathcal{B}_{(k)}^T + \eta^{(t_i)} (U_k^{(t_i+1)} - U_k^{(t_i)}) + U_k^{(t_i+1)} (\mathbf{Y}_k^{(t_i+1)} + (\mathbf{Y}_k^{(t_i+1)})^T)$ , where  $\mathcal{B} = \hat{\mathcal{P}}^{(t_i)} \times_{s+1} U_{s+1}^{(t_i+1)} \cdots \times_{k-1} U_{k-1}^{(t_i+1)}$  and  $\mathcal{A} = \mathcal{Z}^{(t_i+1)} \times_h U_h^{(t_i+1)T} \times_{h-1} \cdots \times_{k+1} U_{k+1}^{(t_i+1)T}$ .

Thus, we have  $(U_k^*)^T U_k^* = I$  and there exists  $\mathbf{Y}_k^*$  such that  $\mathbf{0} = (U_k^* \mathcal{C}_{(k)}^*) \mathcal{B}_{(k)}^{*T} + U_k^* (\mathbf{Y}_k^* + (\mathbf{Y}_k^*)^T)$  if  $i \rightarrow \infty$ , where  $\mathcal{B}^* = \mathcal{Z}^* \times_h (U_h^*)^T \cdots \times_k (U_k^*)^T$  and  $\mathcal{C}^* = \mathcal{Y}^* \times_{s+1} U_{s+1}^* \cdots \times_{k-1} U_{k-1}^*$ . Therefore,  $\mathbf{0} = -\mathcal{F}_{(k)}^* (\mathcal{C}_{(k)}^*)^T + U_k^* (-\mathbf{Y}_k^* + (-\mathbf{Y}_k^*)^T)$  holds, where  $\mathcal{F}^* = \mathcal{Z}^* \times_h (U_h^*)^T \cdots \times_{k+1} (U_{k+1}^*)^T$ .

Besides, from the optimality of  $\mathcal{E}^{(t_i+1)}$ , we have  $\Psi_{\mathbb{I}}(\mathcal{E}^{(t_i+1)}) = \mathbf{0}$  and

$$\Psi_{\mathbb{I}c}(\mu^{(t_i)} (\mathcal{E}^{(t_i+1)} + \mathcal{X}^{(t_i+1)} - \frac{1}{\mu^{(t_i)}} \mathcal{Y}^{(t_i)}) + \eta^{(t_i)} (\mathcal{E}^{(t_i+1)} - \mathcal{E}^{(t_i)})) = \mathbf{0},$$

from which we deduce that both of  $\Psi_{\mathbb{I}}(\mathcal{E}^*) = \mathbf{0}$  and  $\mathbf{0} = \lim_{i \rightarrow \infty} \Psi_{\mathbb{I}c}(\mu^{(t_i)} (\mathcal{E}^{(t_i+1)} - \Psi_{\mathbb{I}}(\mathcal{M}) + \mathcal{X}^{(t_i+1)} - \frac{1}{\mu^{(t_i)}} \mathcal{Y}^{(t_i)})) = -\Psi_{\mathbb{I}c}(\mathcal{Y}^*)$  hold. Furthermore, it is evident that there exists  $\mathcal{W}^*$  such that  $\mathbf{0} = -\Psi_{\mathbb{I}}(\mathcal{Y}^*) + \mathcal{W}^*$ .  $\square$

## 5 PROOF OF LEMMA 1 AND THEOREM 2

**Lemma 1.** For  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_h}$ , the subgradient of  $\|\mathcal{A}\|_{1,\mathcal{U}}$  is given as  $\partial \|\mathcal{A}\|_{1,\mathcal{U}} = \{\mathcal{U}^{-1}(\text{sgn}(\mathcal{U}(\mathcal{A}))) + \mathcal{F} | \Psi_{\mathbb{H}}(\mathcal{F}) = \mathbf{0}, \|\mathcal{F}\|_{\mathcal{U},\infty} \leq 1\}$ , where  $\mathbb{H}$  denotes the support of  $\mathcal{U}(\mathcal{A})$ .

*Proof.* We can get the conclusion by  $\langle \mathcal{U}^{-1}(\text{sgn}(\mathcal{U}(\mathcal{A}))) + \mathcal{F}, \mathcal{A} \rangle = \langle \text{sgn}(\mathcal{U}(\mathcal{A})), \mathcal{U}(\mathcal{A}) \rangle + \langle \mathcal{U}(\mathcal{F}), \mathcal{U}(\mathcal{A}) \rangle = \|\mathcal{A}\|_{1,\mathcal{U}}$  and  $\|\mathcal{U}^{-1}(\text{sgn}(\mathcal{U}(\mathcal{A}))) + \mathcal{F}\|_{\mathcal{U},\infty} = \|\text{sgn}(\mathcal{U}(\mathcal{A})) + \mathcal{U}(\mathcal{F})\|_{\infty} = \max(\|\text{sgn}(\mathcal{U}(\mathcal{A}))\|_{\infty}, \|\mathcal{U}(\mathcal{F})\|_{\infty}) \leq 1$  Watson (1992).  $\square$

**Lemma 2.** If there exists a dual certificate  $\mathcal{G}$  (that satisfy  $\Psi_{\mathbb{I}}(\mathcal{G}) = \mathcal{G}$ ,  $P_{\mathbb{S}}(\mathcal{G}) = \hat{\mathcal{U}}^{-1}(\text{sgn}(\hat{\mathcal{U}}(\mathcal{M})))$ ) and  $\|P_{\mathbb{S}^{\perp}}(\mathcal{G})\|_{\hat{\mathcal{U}},\infty} \leq 1$ ) and any  $\mathcal{H}$  obeying  $\Psi_{\mathbb{I}}(\mathcal{H}) = \mathbf{0}$ , then

$$\|\mathcal{M} + \mathcal{H}\|_{\hat{\mathcal{U}},1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}},1} + (1 - \|P_{\mathbb{S}^{\perp}}(\mathcal{G})\|_{\hat{\mathcal{U}},\infty}) \|P_{\mathbb{S}^{\perp}}(\mathcal{H})\|_{\hat{\mathcal{U}},1}.$$

*Proof.* For any  $\mathcal{Z} \in \partial \|\mathcal{M}\|_{\hat{\mathcal{U}},1}$ , we have  $\|\mathcal{M} + \mathcal{H}\|_{\hat{\mathcal{U}},1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}},1} + \langle \mathcal{Z}, \mathcal{H} \rangle$ . Since  $\mathcal{G} = \hat{\mathcal{U}}^{-1}(\text{sgn}(\hat{\mathcal{U}}(\mathcal{M}))) + P_{\mathbb{S}^{\perp}}(\mathcal{G})$  and  $\mathcal{Z} = \hat{\mathcal{U}}^{-1}(\text{sgn}(\hat{\mathcal{U}}(\mathcal{M}))) + P_{\mathbb{S}^{\perp}}(\mathcal{Z})$ , we obtain  $\|\mathcal{M} + \mathcal{H}\|_{\hat{\mathcal{U}},1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}},1} + \langle \mathcal{G}, \mathcal{H} \rangle + \langle P_{\mathbb{S}^{\perp}}(\mathcal{Z} - \mathcal{G}), \mathcal{H} \rangle$ . Therefore we have  $\|\mathcal{M} + \mathcal{H}\|_{\hat{\mathcal{U}},1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}},1} + \langle P_{\mathbb{S}^{\perp}}(\mathcal{Z} - \mathcal{G}), \mathcal{H} \rangle$ , where  $\langle \mathcal{G}, \mathcal{H} \rangle = \mathbf{0}$  due to  $\Psi_{\mathbb{I}}(\mathcal{H}) = \mathbf{0}$ .

Since  $\|\cdot\|_{\hat{\mathcal{U}},1}$  and  $\|\cdot\|_{\hat{\mathcal{U}},\infty}$  are dual to each other, there exists  $\|\mathcal{Z}_0\|_{\hat{\mathcal{U}},\infty} \leq 1$  such that  $\langle \mathcal{Z}_0, P_{\mathbb{S}^{\perp}}(\mathcal{H}) \rangle = \|P_{\mathbb{S}^{\perp}}(\mathcal{H})\|_{\hat{\mathcal{U}},1}$ . Hence, by selecting a  $\mathcal{Z}$  such that  $P_{\mathbb{S}^{\perp}}(\mathcal{Z}) = P_{\mathbb{S}^{\perp}}(\mathcal{Z}_0)$ , we get  $\langle P_{\mathbb{S}^{\perp}}(\mathcal{Z}), \mathcal{H} \rangle = \|P_{\mathbb{S}^{\perp}}(\mathcal{H})\|_{\hat{\mathcal{U}},1}$ . Therefore, we have  $\langle P_{\mathbb{S}^{\perp}}(\mathcal{Z} - \mathcal{G}), \mathcal{H} \rangle \geq (1 - \|P_{\mathbb{S}^{\perp}}(\mathcal{G})\|_{\hat{\mathcal{U}},\infty}) \|P_{\mathbb{S}^{\perp}}(\mathcal{H})\|_{\hat{\mathcal{U}},1}$  due to  $|\langle P_{\mathbb{S}^{\perp}}(\mathcal{G}), P_{\mathbb{S}^{\perp}}(\mathcal{H}) \rangle| \leq \|P_{\mathbb{S}^{\perp}}(\mathcal{G})\|_{\hat{\mathcal{U}},\infty} \|P_{\mathbb{S}^{\perp}}(\mathcal{H})\|_{\hat{\mathcal{U}},1}$ , thus completed the proof.  $\square$

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{U}_{k_n}^T \mathbf{U}_{k_n} = \mathbf{I}(n=s+1, \dots, h)} \|\mathcal{X} \times_{k_{s+1}} \mathbf{U}_{k_{s+1}} \cdots \times_{k_h} \mathbf{U}_{k_h}\|_{\mathcal{U}, 1} \\ \text{s.t. } \|\Psi_{\mathbb{I}}(\mathcal{M}) - \Psi_{\mathbb{I}}(\mathcal{X})\|_F \leq \delta \end{aligned} \quad (13)$$

**Theorem 2.** *If the dual certificate  $\mathcal{G} = \Psi_{\mathbb{I}} P_{\mathbb{S}} (P_{\mathbb{S}} \Psi_{\mathbb{I}} P_{\mathbb{S}})^{-1} (\hat{\mathcal{U}}^{-1}(\text{sgn}(\hat{\mathcal{U}}(\mathcal{M}))))$  satisfies  $\|P_{\mathbb{S}^\perp}(\mathcal{G})\|_{\hat{\mathcal{U}}, \infty} \leq C_1 < 1$  and  $P_{\mathbb{S}} \Psi_{\mathbb{I}} P_{\mathbb{S}} \succcurlyeq C_2 p \mathcal{I}$ , then we can obtain the following inequality:*

$$\|\mathcal{M} - \hat{\mathcal{X}}\|_F \leq \frac{1}{1 - C_1} \sqrt{\frac{1/C_2 + p}{p}} I_1 I_2 \delta + \delta, \quad (14)$$

where  $\hat{\mathcal{X}}$  is obtained by (13) and  $p$  denotes the sampling rate.

*Proof.* Let  $\mathcal{H} = \hat{\mathcal{X}} - \mathcal{M}$  for brevity. Considering that  $\|\mathcal{H}\|_F = \|\Psi_{\mathbb{I}}(\mathcal{H})\|_F + \|\Psi_{\mathbb{I}^c}(\mathcal{H})\|_F \leq \delta + \|\Psi_{\mathbb{I}^c}(\mathcal{H})\|_F$ , we focus solely on the second term  $\|\Psi_{\mathbb{I}^c}(\mathcal{H})\|_F$  in the following discussion.

Utilizing the triangle inequality and Lemma 2, we obtain  $\|\mathcal{M} + \mathcal{H}\|_{\hat{\mathcal{U}}, 1} \geq \|\mathcal{M} + \Psi_{\mathbb{I}^c}(\mathcal{H})\|_{\hat{\mathcal{U}}, 1} - \|\Psi_{\mathbb{I}}(\mathcal{H})\|_{\hat{\mathcal{U}}, 1}$  and  $\|\mathcal{M} + \Psi_{\mathbb{I}^c}(\mathcal{H})\|_{\hat{\mathcal{U}}, 1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}}, 1} + (1 - \|P_{\mathbb{S}^\perp}(\mathcal{G})\|_{\hat{\mathcal{U}}, \infty}) \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_{\hat{\mathcal{U}}, 1}$ . Consequently, we have  $\|\mathcal{M}\|_{\hat{\mathcal{U}}, 1} \geq \|\mathcal{M} + \mathcal{H}\|_{\hat{\mathcal{U}}, 1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}}, 1} + (1 - \|P_{\mathbb{S}^\perp}(\mathcal{G})\|_{\hat{\mathcal{U}}, \infty}) \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_{\hat{\mathcal{U}}, 1} - \|\Psi_{\mathbb{I}}(\mathcal{H})\|_{\hat{\mathcal{U}}, 1} \geq \|\mathcal{M}\|_{\hat{\mathcal{U}}, 1} + (1 - C_1) \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_{\hat{\mathcal{U}}, 1} - \|\Psi_{\mathbb{I}}(\mathcal{H})\|_{\hat{\mathcal{U}}, 1}$ , which leads to  $\|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F \leq \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_{\hat{\mathcal{U}}, 1} \leq \frac{1}{1 - C_1} \|\Psi_{\mathbb{I}}(\mathcal{H})\|_{\hat{\mathcal{U}}, 1} \leq \frac{\sqrt{I_1 I_2}}{1 - C_1} \|\Psi_{\mathbb{I}}(\mathcal{H})\|_F \leq \frac{\sqrt{I_1 I_2}}{1 - C_1} \delta$ .

Additionally, due to  $P_{\mathbb{S}} \Psi_{\mathbb{I}} P_{\mathbb{S}} \succcurlyeq C_2 p \mathcal{I}$ , we find  $\|\Psi_{\mathbb{I}}(P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H})))\|_F^2 = \langle P_{\mathbb{S}} \Psi_{\mathbb{I}} P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H})), P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H})) \rangle \geq C_2 p \|P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F^2$ . Moreover, because of  $\Psi_{\mathbb{I}}(P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))) + \Psi_{\mathbb{I}}(P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H}))) = \mathbf{0}$ , we get  $C_2 p \|P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F^2 \leq \|\Psi_{\mathbb{I}}(P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H})))\|_F^2 = \|\Psi_{\mathbb{I}}(P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H})))\|_F^2 \leq \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F^2$ .

Consequently, we have  $\|\Psi_{\mathbb{I}^c}(\mathcal{H})\|_F^2 = \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F^2 + \|P_{\mathbb{S}}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F^2 \leq (\frac{1}{C_2 p} + 1) \|P_{\mathbb{S}^\perp}(\Psi_{\mathbb{I}^c}(\mathcal{H}))\|_F^2 \leq (\frac{1}{C_2 p} + 1) \frac{I_1 I_2}{(1 - C_1)^2} \delta^2$ , and thus completed the proof.  $\square$

## 6 STABLE TC-SL

Similarly, we can establish stable TC-SL based on the given  $\{\hat{\mathcal{U}}_{k_n}\}_{n=3}^s$ :

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{U}_{k_n}^T \mathbf{U}_{k_n} = \mathbf{I}(n=s+1, \dots, h)} \|\mathcal{X} \times_{k_{s+1}} \mathbf{U}_{k_{s+1}} \cdots \times_{k_h} \mathbf{U}_{k_h}\|_{*, \mathcal{U}}^{(k_1, k_2)} \\ \text{s.t. } \|\Psi_{\mathbb{I}}(\mathcal{M}) - \Psi_{\mathbb{I}}(\mathcal{X})\|_F \leq \delta. \end{aligned} \quad (15)$$

Before proving the stable recovery property of (15), we need to introduce the definition of the tensor product, which is a direct generalization from high order tensor product defined in Qin et al. (2022).

**Definition 3.** (tensor product for given  $(k_1, k_2)$  and  $\mathcal{U}$ ) For an  $h$ -order tensor  $\mathcal{A} \in \mathbb{R}^{I_{k_1} \times L \times \cdots \times I_{k_h}}$  and  $\mathcal{B} \in \mathbb{R}^{L \times I_{k_2} \times \cdots \times I_{k_h}}$ , the tensor product of  $\mathcal{A}$  and  $\mathcal{B}$  is defined as  $\mathcal{A} * \mathcal{B} = \mathcal{U}^{-1}(\mathcal{U}(\mathcal{A}) \odot_f \mathcal{U}(\mathcal{B}))$ , where  $[\tilde{\mathcal{A}} \odot_f \tilde{\mathcal{B}}]_{:, :, i_{k_3}, i_{k_4}, \dots, i_{k_h}} = [\tilde{\mathcal{A}}]_{:, :, i_{k_3}, i_{k_4}, \dots, i_{k_h}} [\tilde{\mathcal{B}}]_{:, :, i_{k_3}, i_{k_4}, \dots, i_{k_h}}$ .

Let  $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$  be t-SVD of  $\mathcal{A}$  by using tensor product given in the Definition 3, where  $\mathcal{V}^T$  is defined by  $[\mathcal{U}(\mathcal{V}^T)]_{:, :, i_{k_3}, i_{k_4}, \dots, i_{k_h}} = [\mathcal{U}(\mathcal{V})]_{:, :, i_{k_3}, i_{k_4}, \dots, i_{k_h}}^T$  for all  $(i_{k_3}, i_{k_4}, \dots, i_{k_h})$ . For simplicity, we'll consider the case of  $(k_1, k_2, \dots, k_h) = (1, 2, \dots, h)$  and use  $\|\cdot\|_{*, \mathcal{U}}$  and  $\|\cdot\|_{2, \mathcal{U}}$  to denote  $\|\cdot\|_{*, \mathcal{U}}^{(1, 2)}$  and  $\|\cdot\|_{2, \mathcal{U}}^{(1, 2)}$ , respectively.

**Lemma 3.** For tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_h}$  with  $\text{rank}_{(1, 2)}(\mathcal{U}(\mathcal{A})) = r$ , if its skinny t-SVD is  $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ , then the subgradient of  $\|\mathcal{A}\|_{*, \mathcal{U}}$  can be given as  $\partial \|\mathcal{A}\|_{*, \mathcal{U}} = \{\mathcal{U} * \mathcal{V}^T + \mathcal{W} | \mathcal{U}^T * \mathcal{W} = \mathbf{0}, \mathcal{W} * \mathcal{V} = \mathbf{0}, \|\mathcal{W}\|_{2, \mathcal{U}} \leq 1\}$ .

Table 1: Comparing various methods on the five video segments at a sampling rate  $p = 0.3$ .

Video	TNN-DCT	TNN-DFT	SNN	KBR	WSTNN	HTNN-DCT	TC-SL	TC-U1
run 9 <sup>th</sup>	25.77	25.73	22.53	27.48	30.54	28.35	30.63	<b>32.79</b>
run 39 <sup>th</sup>	30.66	30.60	29.24	38.03	34.74	34.05	35.01	<b>40.39</b>
run 40 <sup>th</sup>	28.83	28.80	26.13	33.1	32.59	31.73	33.35	<b>36.06</b>
run 42 <sup>th</sup>	27.72	27.86	24.48	31.75	32.08	30.63	31.87	<b>36.88</b>
run 108 <sup>th</sup>	31.64	31.55	29.83	34.13	34.13	33.72	35.57	<b>36.96</b>
Average	28.92	28.91	26.44	32.90	32.82	31.70	33.29	<b>36.62</b>

*Proof.* We can obtain the conclusion by  $\langle \mathcal{U} * \mathcal{V}^T + \mathcal{W}, \mathcal{A} \rangle = \langle \mathcal{U} * \mathcal{V}^T, \mathcal{U} * \mathcal{S} * \mathcal{V}^T \rangle + \langle \mathcal{W}, \mathcal{U} * \mathcal{S} * \mathcal{V}^T \rangle = \langle \mathcal{I}, \mathcal{S} \rangle = \|\mathcal{A}\|_{*,\mathcal{U}}$  and  $\|\mathcal{U} * \mathcal{V}^T + \mathcal{W}\|_{2,\mathcal{U}} \leq 1$  Watson (1992).  $\square$

Suppose  $\mathcal{M} = \mathcal{U}_0 * \mathcal{S}_0 * \mathcal{V}_0^T$  is the skinny t-SVD of  $\mathcal{M}$ . We define  $\mathbb{T} = \{\mathcal{U}_0 * \mathcal{Y}^T + \mathcal{W} * \mathcal{V}_0^T, \mathcal{Y}, \mathcal{W} \in \mathbb{R}^{I_1 \times r \times I_3 \times \dots \times I_h}\}$ ,  $P_{\mathbb{T}}$  is the projections onto  $\mathbb{T}$ , and  $\mathbb{T}^\perp$  is the orthogonal complement of  $\mathbb{T}$ . Considering  $(\hat{\mathcal{X}}, \{\hat{U}_k\}_{k=1}^h)$  as the result obtained by (15), we define  $\hat{\mathcal{U}}(\mathcal{A}) = \mathcal{A} \times_1 \hat{U}_1 \times_2 \dots \times_h \hat{U}_h$ . By the property of subgradient  $\partial \|\cdot\|_{*,\hat{\mathcal{U}}}$  and the duality between  $\|\mathcal{W}\|_{2,\hat{\mathcal{U}}}$  and  $\|\mathcal{W}\|_{*,\hat{\mathcal{U}}}$ , we can get the following results.

**Lemma 4.** *If there exists a dual certificate  $\mathcal{G}$  (that satisfy  $\Psi_{\mathbb{T}}(\mathcal{G}) = \mathcal{G}$ ,  $P_{\mathbb{T}}(\mathcal{G}) = \mathcal{U}_0 * \mathcal{V}_0^T$  and  $\|P_{\mathbb{T}^\perp}(\mathcal{G})\|_{2,\hat{\mathcal{U}}} \leq 1$ ), we have*

$$\|\mathcal{M} + \mathcal{H}\|_{*,\hat{\mathcal{U}}} \geq \|\mathcal{M}\|_{*,\hat{\mathcal{U}}} + (1 - \|P_{\mathbb{T}^\perp}(\mathcal{G})\|_{2,\hat{\mathcal{U}}})\|P_{\mathbb{T}^\perp}(\mathcal{H})\|_{*,\hat{\mathcal{U}}}$$

for any  $\mathcal{H}$  obeying  $\Psi_{\mathbb{T}}(\mathcal{H}) = 0$ .

**Theorem 3.** *If the dual certificate  $\mathcal{G} = \Psi_{\mathbb{T}}P_{\mathbb{T}}(P_{\mathbb{T}}\Psi_{\mathbb{T}}P_{\mathbb{T}})^{-1}(\mathcal{U}_0 * \mathcal{V}_0^T)$  satisfies  $\|P_{\mathbb{T}^\perp}(\mathcal{G})\|_{2,\hat{\mathcal{U}}} \leq C_1 < 1$  and  $P_{\mathbb{T}}\Psi_{\mathbb{T}}P_{\mathbb{T}} \succcurlyeq C_2 p \mathcal{I}$ , then we have*

$$\|\mathcal{M} - \hat{\mathcal{X}}\|_F \leq \frac{1}{1 - C_1} \sqrt{\frac{1/C_2 + p}{p} \min(I_1, I_2) \delta} + \delta, \quad (16)$$

where  $\hat{\mathcal{X}}$  is obtained by (15) and  $p$  is the sampling rate.

## 7 COLOR VIDEO INPAINTING

We randomly selected five color video segments with the most rapidly changing frames from category ‘run’ of the *HMDB51*, including run 9<sup>th</sup>, run 39<sup>th</sup>, run 40<sup>th</sup>, run 42<sup>th</sup>, and run 108<sup>th</sup>, and evaluated all tensor completion methods on the selected video segments, where run  $x^{\text{th}}$  is used to represent the  $x$ -th video in the category ‘run’. We present the PSNR values of all methods on the five video segments in Table 1. The results in the table show a significant improvement achieved by our methods (TC-SL and TC-U1) for color video inpainting. The PSNR results obtained by TC-U1 outperform the third-best method (the second-best method is TC-SL) by more than 3.5 dB on average. This substantial improvement showcased by TC-U1 in color video inpainting, as reflected in the higher PSNR values, provides strong evidence for its effectiveness in high-order tensor completion, particularly in scenarios involving non-smooth changes between tensor slices.

## REFERENCES

- Frank H Clarke. Nonsmooth analysis and optimization. In *Proceedings of the international congress of mathematicians*, volume 5, pp. 847–853. Citeseer, 1983.
- Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.
- Maryam Fazel Sarjoui. Matrix rank minimization with applications. 2002.



- 
- Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.
- Jean Baptiste Hiriarturruty and Claude Lemaréchal. Convex analysis and minimization algorithms ii: Advanced theory and bundle methods. *Grundlehren Der Mathematischen Wissenschaften*, 1(1185):150–159, 1993.
- Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.
- Frank L Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, 7(1-4):39–79, 1928.
- Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009a. doi: 10.1137/07070111X.
- Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009b.
- Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 35(1): 208–220, 2013.
- Canyi Lu, Jiashi Feng, Zhouchen Lin, and Shuicheng Yan. Exact low tubal rank tensor recovery from gaussian measurements. *arXiv preprint arXiv:1806.02511*, 2018.
- Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):925–938, 2019a.
- Canyi Lu, Xi Peng, and Yunchao Wei. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5996–6004, 2019b.
- Wenjin Qin, Hailin Wang, Feng Zhang, Jianjun Wang, Xin Luo, and Tingwen Huang. Low-rank high-order tensor completion with applications in visual data. *IEEE Transactions on Image Processing*, 31:2433–2448, 2022. doi: 10.1109/TIP.2022.3155949.
- Ledyard R Tucker. Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change*, 15(122-137):3, 1963.
- Hailin Wang, Jiangjun Peng, Wenjin Qin, Jianjun Wang, and Deyu Meng. Guaranteed tensor recovery fused low-rankness and smoothness, 2023.
- G Alistair Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992.
- Xiaoqin Zhang, Zhengyuan Zhou, Di Wang, and Yi Ma. Hybrid singular value thresholding for tensor completion. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014a.
- Xiaoqin Zhang, Jingjing Zheng, Li Zhao, Zhengyuan Zhou, and Zhouchen Lin. Tensor recovery with weighted tensor average rank. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-svd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3842–3849, 2014b.
- Jingjing Zheng, Xiaoqin Zhang, Wenzhe Wang, and Xianta Jiang. Handling slice permutations variability in tensor recovery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3): 3499–3507, 2022.

---

Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Tai-Xiang Jiang, Teng-Yu Ji, and Tian-Hui Ma. Tensor n-tubal rank and its convex relaxation for low-rank tensor recovery. *Information Sciences*, 532: 170–189, 2020. ISSN 0020-0255. doi: <https://doi.org/10.1016/j.ins.2020.05.005>. URL <https://www.sciencedirect.com/science/article/pii/S0020025520303923>.

Pan Zhou, Canyi Lu, Zhouchen Lin, and Chao Zhang. Tensor factorization for low-rank tensor completion. *IEEE Transactions on Image Processing*, 27(3):1152–1163, 2017.