# BAYESIAN-DRIVEN LEARNING OF A NEW WEIGHTED TENSOR NORM FOR TENSOR RECOVERY

**Jingjing Zheng**
Department of Mathematics
The University of British Columbia
British Columbia, Canada
jjzheng@math.ubc.ca

**Yankai Cao***
Department of Chemical and Biological Engineering
The University of British Columbia
British Columbia, Canada
yankai.cao@ubc.ca

## ABSTRACT

This study addresses the performance limitations of t-SVD-based tensor recovery caused by non-smooth changes and imbalanced low-rankness in tensor data. We introduce a novel bilevel tensor completion model, integrating the learning of a data-dependent weighted tensor norm within the tensor completion framework as an upper-level problem. We treat the optimization of the bilevel problem as a black-box problem, employing Bayesian Optimization (BO) for efficient learning of the proposed tensor norm. Numerical experiments demonstrated the superior performance of our proposed method compared to state-of-the-art methods in tensor completion. The code of our method is available at https://github.com/jzheng20/TR-BO.git.

## 1 INTRODUCTION

Currently, the tensor recovery can classified into three main categories: CP decomposition-based methods (Hitchcock, 1927; 1928), Tucker decomposition-based methods (Tucker, 1963; Liu et al., 2013; Xie et al., 2017), and tensor product (t-product)-based methods (Lu et al., 2019b; 2018; Qin et al., 2022; Zheng et al., 2022; Wang et al., 2023; Zhang et al., 2022; Zhou et al., 2017). Due to the effectiveness of t-product-based methods in image and video processing, this work mainly focus on the developing of the current t-product-based methods.

Different with the first two defining ways for tensor rank, the t-product-based methods are looking for slice-wise tensor decomposition of the resulting tensor obtained by performing a fixed transform along the certain dimension of given tensor data. For example, Kilmer & Martin (2011) given Discrete Fourier Transform (DFT)-based t-product defined on the face product of the transformed tensors by DFT. After determining the t-product, the tensor tubal rank (or tensor average rank) (Lu et al., 2019a) for characterizing the internal correlation of tensor data can be naturally given by counting the number of non-zero singular tubal (or values) obtained by tensor Singular Value Decomposition (t-SVD). To study the low-rankness across different dimensions of higher order tensor data, Zheng et al. (2020) have proposed a new tensor norm by using the Weighted Sum of Tensor Nuclear Norm of all mode-$k_1k_2$ unfolding tensors (WSTNN). However, $\binom{h}{2}$ different unfolding tensors are considered in WSTNN for $h$-order tensor data and it leads to a difficult setting for the weight parameters, with the increasing $h$, especially when tensor data has imbalanced low-rankness across its different dimensions. Besides, since fixed transform are used in these methods, these tensor product-based methods suffer from significant performance degradation when dealing with tensor data that exhibits fast-changing or non-continuous variations as shown in the Fig. 1 of the appendix. The contribution of this work is to give a new tensor recovery framework based on a novel weighted tensor norm from bayesian-driven learning to solve the above issues, and thus can recover the low-rank component more effectively.

## 2 METHODOLOGY

In this section, we propose a novel weighted tensor norm formulized as

$$\|\boldsymbol{\mathcal{X}}\|_{*,\mathcal{U}}^{\boldsymbol{w}} = \sum_{1\le i<j\le h} w_{i,j}\|\boldsymbol{\mathcal{X}} \times_{k_1} \hat{\boldsymbol{U}}_{k_1} \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s} \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}} \cdots \boldsymbol{U}_{i,j,k_h}\|_*^{(i,j)}, \tag{1}$$

*Corresponding Author

Table 1: Comparing the PSNR results by all methods on *Berkeley Segmentation Dataset* (*BSD*) (Martin et al., 2001) at different sampling rates $c$. 'Ours' and 'Ours-BO' represents the proposed norm with the equal weights and weights learned from BO, respectively.

| Sampling Rate $c$ | TNN-DCT | SNN | KBR | WSTNN | HTNN-DCT | Ours | Ours-BO |
|---|---|---|---|---|---|---|---|
| 0.3 | 23.25 | 21.86 | 25.45 | 25.75 | 25.21 | 27.87 | **28.33** |
| 0.5 | 27.25 | 25.50 | 31.57 | 31.07 | 30.72 | 33.35 | **33.82** |
| 0.7 | 32.04 | 29.84 | 38.81 | 37.11 | 38.22 | 40.47 | **40.81** |

which aims to study the low-rank characteristics of an h-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_h}$ across its various dimensions. Here, $\|\boldsymbol{\mathcal{A}}\|_*^{(k_1,k_2)} = \sum_{1 \leq i_{k_n} \leq I_{k_n}(n=3,4,\cdots,h)} \|[\boldsymbol{\mathcal{A}}]_{i_1,\cdots,i_{k_1-1},:,i_{k_1+1},\cdots,i_{k_2-1},:,i_{k_2+1},\cdots,i_{k_h}}\|_*$, and $\times_n$ is mode-n product. We have proposed

$$[\boldsymbol{\mathcal{X}}^*, \boldsymbol{U}_{i,j,k_n}^*] = \underset{\boldsymbol{\mathcal{X}}, \boldsymbol{U}_{i,j,k_n}^T \boldsymbol{U}_{i,j,k_n} = \boldsymbol{I}}{\arg\min} \|\boldsymbol{\mathcal{X}}\|_{*,\mathcal{U}}^{\boldsymbol{w}} \qquad s.t.\ \mathbf{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathbf{P}_\Omega(\boldsymbol{\mathcal{M}}) \tag{2}$$

for a better tensor recovery and the following bilevel tensor completion model for the learning of the proposed weighted norm:

$$\begin{aligned}[\boldsymbol{w}^{\mathrm{bo}}, \boldsymbol{U}_{i,j,k_n}^{\mathrm{bo}}] = & \underset{\boldsymbol{w}, \boldsymbol{U}_{i,j,k_n}^{\mathrm{bo}} = \boldsymbol{U}_{i,j,k_n}^*}{\arg\min} \|\mathbf{P}_\Phi(\boldsymbol{\mathcal{Z}}^*) - \mathbf{P}_\Phi(\boldsymbol{\mathcal{M}})\|_F \\ & s.t.\ [\boldsymbol{\mathcal{Z}}^*, \boldsymbol{U}_{i,j,k_n}^*] = \underset{\boldsymbol{\mathcal{Z}}, \boldsymbol{U}_{i,j,k_n}^T \boldsymbol{U}_{i,j,k_n} = \boldsymbol{I}}{\arg\min} \|\boldsymbol{\mathcal{Z}}\|_{*,\mathcal{U}}^{\boldsymbol{w}} \qquad s.t.\ \mathbf{P}_{\Phi^c \cap \Omega}(\boldsymbol{\mathcal{Z}}) = \mathbf{P}_{\Phi^c \cap \Omega}(\boldsymbol{\mathcal{M}}),\end{aligned} \tag{3}$$

where $\Omega$ denotes the set of locations of the observed elements in the tensor $\boldsymbol{\mathcal{M}}$, and $\Phi \subset \Omega$. Once we get the learned weighted norm based on $\boldsymbol{U}_{i,j,k_n}^{\mathrm{bo}}(n = s+1, \cdots, h)$ and $\boldsymbol{w}^{\mathrm{bo}}$, we can obtain a better estimation $\boldsymbol{\mathcal{X}}^*$ for $\boldsymbol{\mathcal{M}}$ by minimizing the proposed weighted norm $\|\boldsymbol{\mathcal{X}}\|_{*,\mathcal{U}}^{\boldsymbol{w}^{\mathrm{bo}}}$ under the constraint of $\mathbf{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathbf{P}_\Omega(\boldsymbol{\mathcal{M}})$. [1] From convergence guarantee for the solving $\boldsymbol{\mathcal{Z}}^*$ by using APMM (Zheng et al., 2023) that is given in the appendix, we can provides an approximate representation for $\boldsymbol{\mathcal{Z}}^*$ as $f_{\boldsymbol{w}}^{(N)} \circ \cdots \circ f_{\boldsymbol{w}}^{(2)} \circ f_{\boldsymbol{w}}^{(1)}(\mathbf{P}_{\Phi^c \cap \Omega}(\boldsymbol{\mathcal{M}}))$, where $f_{\boldsymbol{w}}^{(t)}(\cdot)$ denotes the $t$-th iteration of the optimization for solving $\boldsymbol{\mathcal{Z}}^{(t)}$ and it depend on the choice of $\boldsymbol{w}$. Consequently, we turn to minimize the approximation of the upper-level objective function, $F_{\boldsymbol{w}} = \|\mathbf{P}_\Phi(f_{\boldsymbol{w}}^{(N)} \circ \cdots \circ f_{\boldsymbol{w}}^{(2)} \circ f_{\boldsymbol{w}}^{(1)}(\mathbf{P}_{\Phi^c \cap \Omega}(\boldsymbol{\mathcal{M}}))) - \mathbf{P}_\Phi(\boldsymbol{\mathcal{M}})\|_F$, using BO. We assume that $F_{\boldsymbol{w}}$ follows a Gaussian Process $\mathcal{GP}(m(\boldsymbol{w}), k(\boldsymbol{w}, \boldsymbol{w}'))$, and employ the knowledge gradient (Garnett, 2023) as our acquisition function.

## 3 Experiments

In this section, we compared the proposed methods with five state-of-the-art methods, including TNN-DCT (Lu et al., 2019b), SNN (Liu et al., 2013), KBR (Xie et al., 2017), WSTNN (Zheng et al., 2020), and HTNN-DCT (Qin et al., 2022), on image sequence inpainting to demonstrate the effectiveness of the proposed methods in tensor completion. The Peak Signal-To-Noise Ratio (PSNR) results for all methods are presented in the Table 1[2]. From the table, we can see that our methods (Ours and Ours-BO) have achieved the best performance for all cases. Particularly, the results achieved by Ours-BO exhibit an improvement of approximately 0.5 dB compared to those attained by Ours. Furthermore, when compared to other methods, our approaches demonstrate a superiority of at least 2 dB across all cases.

## 4 Conclusion

In this work, we have proposed a new data-dependent weighted tensor norm learned by BO for handling the case when tensor data is imbalanced low-rankness and non-continuous changing. The experiments demonstrates the superior performance of the proposed method. It worth noting that, beyond tensor completion, the proposed framework can be applied to other tensor recovery model, offering numerous possibilities for future research directions.

[1] In practice, we sample the slices in $\mathbf{P}_{\Phi^c \cap \Omega}(\boldsymbol{\mathcal{M}})$ to construct a smaller tensor for a more efficient learning of $\boldsymbol{w}$, and we have detailed the algorithm for that case in the appendix.

[2] Following Lu et al. (2019a), we randomly selected 50 color images.

ACKNOWLEDGEMENTS

Yankai Cao acknowledges funding from New Frontiers in Research Fund and the discovery program of the Natural Science and Engineering Research Council of Canada. The authors also gratefully acknowledge the computing resources and services provided by SciNet (www.scinethpc.ca), the Digital Research Alliance of Canada (www.alliancecan.ca), and Advanced Research Computing at the University of British Columbia.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

## REFERENCES

Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

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.

Misha E Kilmer and Carla D Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.

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.

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision.*, volume 2, pp. 416–423, 2001.

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.

Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1888–1902, 2017.

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.


(a) Imbalanced low-rankness across different dimensions (b) Image sequence obtained by concatenating different scenes (c) Rapidly changing content between frames (d) Disordered image samples for classification task

Figure 1: Imbalanced low-rankness and fast-changing in tensor data

---

**Algorithm 1:** BO for the learning of weighted tensor norm

---

**Input:** $\mathbf{P}_\Omega(\boldsymbol{\mathcal{M}})$, $\Phi \subset \Omega$, initial dataset $\mathbb{D}$, $\hat{\boldsymbol{U}}_{k_n}$ for $1 \leq n \leq s$.
**Output:** $\boldsymbol{w}^{\text{bo}}$, $\{\boldsymbol{U}^*_{i,j,k_{n_0}}\}_{n_0=s+1}^{h}$ for $1 \leq i < j \leq h$, and $\boldsymbol{\mathcal{X}}^*$.
1. Sampling the slices in $\mathbf{P}_\Phi(\boldsymbol{\mathcal{M}})$ to obtain $\mathbf{P}_\Phi(\boldsymbol{\mathcal{M}}_{\text{small}})$
2.**Repeat**
3. $\boldsymbol{w} \leftarrow \text{POLICY}(\mathbb{D})$
4. $[\boldsymbol{\mathcal{Z}}^*_{\text{small}}, \boldsymbol{U}^*_{i,j,k_n}] = \arg\min_{\boldsymbol{\mathcal{Z}}_{\text{small}}, \boldsymbol{U}^T_{i,j,k_n}\boldsymbol{U}_{i,j,k_n}=\boldsymbol{I}} \|\boldsymbol{\mathcal{Z}}_{\text{samll}}\|^{\boldsymbol{w}}_{*,\mathcal{U}},\ s.t.\ \mathbf{P}_{\Phi^c\cap\Omega}(\boldsymbol{\mathcal{M}}_{\text{small}}) = \mathbf{P}_{\Phi^c\cap\Omega}(\boldsymbol{\mathcal{Z}}_{\text{samll}})$
5. $F_{\boldsymbol{w}} \leftarrow \text{MSE}(\mathbf{P}_\Phi(\boldsymbol{\mathcal{Z}}^*_{\text{small}}), \mathbf{P}_\Phi(\boldsymbol{\mathcal{M}}_{\text{small}}))$, $\mathbb{D} \leftarrow \mathbb{D} \cup \{(\boldsymbol{w}, F_{\text{mse}})\}$
7.**Until** termination reached
8. Obtain $\boldsymbol{\mathcal{X}}^*$ and $\{\boldsymbol{U}^*_{i,j,k_{n_0}}\}_{n_0=s+1}^{h}$ for $1 \leq i < j \leq h$ by solving
$\arg\min_{\boldsymbol{\mathcal{X}}, \boldsymbol{U}_{i,j,k_n}} \|\boldsymbol{\mathcal{X}}\|^{\boldsymbol{w}^{\text{ob}}}_{*,\mathcal{U}} \quad s.t.\ \mathbf{P}_\Omega(\boldsymbol{\mathcal{M}}) = \mathbf{P}_\Omega(\boldsymbol{\mathcal{X}}),\ \boldsymbol{U}^T_{i,j,k_n}\boldsymbol{U}_{i,j,k_n} = \boldsymbol{I}(n = s+1, \cdots, h,\ 1 \leq i < j \leq h).$

---

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.

Jingjing Zheng, Wanglong Lu, Wenzhe Wang, Yankai Cao, Xiaoqin Zhang, and Xianta Jiang. Handling the non-smooth challenge in tensor SVD: A multi-objective tensor recovery framework, 2023.

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.

# A Appendix

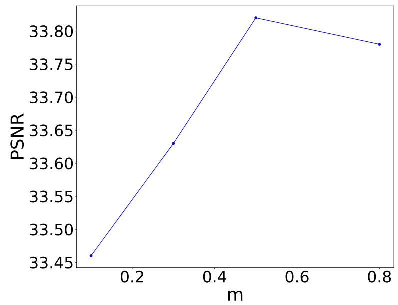
## A.1 The Algorithm for learning the weighted tensor norm by BO (see Algorithm 1 )

## A.2 Parameters analysis

To explore the influence of the choice of $\Phi$, we study the relation between $m = |\Phi^c \cap \Omega|/|\Omega|$ and the recovery performance of the proposed method (Ours-BO). As shown in the Fig. A.2, the performance of Ours-BO have achieved the best when $m = 0.5$.

## A.3 Some properties of the proposed norm

For the convenience of discussion, we define $\|\boldsymbol{\mathcal{A}}\|^{(i,j)}_{*,\mathcal{U}_{i,j}} = \|\mathcal{U}_{i,j}(\boldsymbol{\mathcal{A}})\|^{(i,j)}_* = \|\boldsymbol{\mathcal{A}} \times_{k_1} \hat{\boldsymbol{U}}_{k_1} \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s} \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}} \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}\|^{(i,j)}_*$, and we can get the following results by the convexity of $\|\cdot\|^{(i,j)}_{*,\mathcal{U}_{i,j}}$ and the linearity of the subgradient.




**Property 1.** *The proposed norm defined as* $\|\mathcal{A}\|_{*,\mathcal{U}}^{\boldsymbol{w}} = \sum_{1\le i<j\le h} w_{i,j}\|\mathcal{A}\|_{*,\mathcal{U}_{i,j}}^{(i,j)}$ *for tensor* $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_h}$ *is convex if* $w_{i,j} \ge 0$ *for any* $1 \le i < j \le h$.

**Definition 1.** *(tensor product under mode* $(k_1, k_2)$ *for given* $\mathcal{U}$*) For an h-order tensor* $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots I_{k_2-1}\times L\times I_{k_2+1}\times\cdots\times I_h}$ *and* $\mathcal{B} \in \mathbb{R}^{I_1\times I_2\cdots\times I_{k_1-1}\times L\times I_{k_1+1}\cdots\times I_h}$, *the tensor product of* $\mathcal{A}$ *and* $\mathcal{B}$ *under mode* $(k_1, k_2)$ *is defined as* $\mathcal{A} *_{k_1,k_2} \mathcal{B} = \mathcal{U}^{-1}(\mathcal{U}(\mathcal{A}) \odot_{\mathrm{slice}_{k_1,k_2}} \mathcal{U}(\mathcal{B}))$, *where* $[\bar{\mathcal{A}} \odot_{\mathrm{slice}_{k_1,k_2}} \bar{\mathcal{B}}]_{i_1,\cdots,i_{k_1-1},:,i_{k_1+1},\cdots,i_{k_2-1},:,i_{k_2+1},\cdots,i_{k_h}} = [\bar{\mathcal{A}}]_{i_1,\cdots,i_{k_1-1},:,i_{k_1+1},\cdots,i_{k_2-1},:,i_{k_2+1},\cdots,i_{k_h}}[\bar{\mathcal{B}}]_{i_1,\cdots,i_{k_1-1},:,i_{k_1+1},\cdots,i_{k_2-1},:,i_{k_2+1},\cdots,i_{k_h}}$.

**Lemma 1.** *For tensor* $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_h}$, *if its skinny t-SVD under mode* $(i,j)$ *is* $\mathcal{A} = \mathcal{U}_{i,j} *_{i,j} \mathcal{S}_{i,j} *_{i,j} \mathcal{V}_{i,j}^T$, *then the subgradient of* $\|\mathcal{A}\|_{*,\mathcal{U}}^{\boldsymbol{w}} = \sum_{1\le i<j\le h} w_{i,j}\|\mathcal{A}\|_{*,\mathcal{U}_{i,j}}^{(i,j)}$ *can be given as* $\partial\|\mathcal{A}\|_{*,\mathcal{U}}^{\boldsymbol{w}} = \{\sum_{1\le i<j\le h} w_{i,j}(\mathcal{U}_{i,j} *_{i,j} \mathcal{V}_{i,j}^T + \mathcal{W}_{i,j}) | \mathcal{U}_{i,j}^T *_{i,j} \mathcal{W}_{i,j} = \mathbf{0}, \mathcal{W}_{i,j} *_{i,j} \mathcal{V}_{i,j} = \mathbf{0}, \max_{1\le i<j\le h}\|\mathcal{W}_{i,j}\|_{2,\mathcal{U}_{i,j}}^{(i,j)} \le 1\}$, *where* $\|\mathcal{W}_{i,j}\|_{2,\mathcal{U}_{i,j}}^{(i,j)}$ *is defined as* $\|\mathcal{W}_{i,j}\|_{2,\mathcal{U}_{i,j}}^{(k_1,k_2)} = \max_{i_1,i_2,\cdots,i_{k_1-1},i_{k_1+1},\cdots,i_{k_2-1},i_{k_2+1},\cdots,i_h}\|[\mathcal{U}_{i,j}(\mathcal{W}_{i,j})]_{i_1,\cdots,i_{k_1-1},:,i_{k_1+1},\cdots,i_{k_2-1},:,i_{k_2+1},\cdots,i_{k_h}}\|_2$.

### A.4 Optimization for equation 4 and the corresponding convergence analysis

In this subsection, we are going to solve equation 4 by using APMM Zheng et al. (2023).

$$\begin{aligned}\mathop{\arg\min}_{\mathcal{X},\boldsymbol{U}_{i,j,k_n}} \|\mathcal{X}\|_{*,\mathcal{U}}^{\boldsymbol{w}} \qquad & s.t.\ \mathbf{P}_\Omega(\mathcal{M}) = \mathbf{P}_\Omega(\mathcal{X}),\\ & \boldsymbol{U}_{i,j,k_n}^T\boldsymbol{U}_{i,j,k_n} = \boldsymbol{I}(n = s+1,\cdots,h,\ 1\le i<j\le h). \end{aligned} \tag{4}$$

For easily solving of equation 4, we introduce auxiliary variables $\mathcal{E} \in \mathbb{E} = \{\mathcal{E}|\mathbf{P}_\Omega(\mathcal{E}) = \mathbf{0}\}$ and $\mathcal{Z}_{i,j} = \mathcal{X} \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}} \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}$ for $1 \le i < j \le h$, and thus obtain the following problem:

$$\begin{aligned}\mathop{\arg\min}_{\mathcal{X},\boldsymbol{U}_{i,j,k_n}^T\boldsymbol{U}_{i,j,k_n}=\boldsymbol{I},\mathcal{E}\in\mathbb{E}} & \sum_{1\le i<j\le h} w_{i,j}\|\mathcal{Z}_{i,j}\|_*^{(i,j)}\\ & s.t.\ \mathbf{P}_\Omega(\mathcal{M}) = \mathcal{X} + \mathcal{E}\\ & \mathcal{X} = \mathcal{Z}_{i,j} \times_{k_1} \hat{\boldsymbol{U}}_{k_1}^T \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s}^T \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}}^T \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}^T\\ & (n = s+1,\cdots,h,\ 1\le i<j\le h),\end{aligned} \tag{5}$$

where the Lagrangian function of equation 5 is given as follows

$$\begin{aligned}&\mathcal{L}(\mathcal{X},\mathcal{E},\boldsymbol{U}_{i,j,k},\mathcal{Z}_{i,j},\boldsymbol{\Lambda}_{i,j},\boldsymbol{\Lambda},\mu,\mu_{ij}) = \sum_{1\le i\le j\le h} w_{i,j}\|\mathcal{Z}_{i,j}\|_*^{(i,j)} + \frac{\mu}{2}\|\mathbf{P}_\Omega(\mathcal{M}) - \mathcal{X} - \mathcal{E}\|_F^2\\ &\langle\mathbf{P}_\Omega(\mathcal{M}) - \mathcal{X} - \mathcal{E},\boldsymbol{\Lambda}\rangle + \sum_{1\le i<j\le h}\Big(\frac{\mu_{ij}}{2}\|\mathcal{Z}_{i,j} \times_{k_1} \hat{\boldsymbol{U}}_{k_1}^T \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s}^T \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}}^T \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}^T - \mathcal{X}\|_F^2\\ &+ \langle\mathcal{Z}_{i,j} \times_{k_1} \hat{\boldsymbol{U}}_{k_1}^T \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s}^T \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}}^T \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}^T - \mathcal{X},\boldsymbol{\Lambda}_{ij}\rangle\Big).\end{aligned} \tag{6}$$

According to the framework of APMM, the above optimization problem equation 5 can be iteratively solved by minimizing the Lagrangian function as presented in the Algorithm 2. The convergence of the algorithm is guaranteed by the Theorem 1.

**Algorithm 2:** APMM-based Iterative Solver to equation 5

**Input:** $\mathbf{P}_\Omega(\boldsymbol{\mathcal{M}})$, $\{\boldsymbol{U}_{i,j,k_n}^{(0)}\}_{n=s+1}^h$, $\{\hat{\boldsymbol{U}}_{k_n}\}_{n=1}^s$, $\boldsymbol{\mathcal{E}}^{(0)}$, $\boldsymbol{\mathcal{Y}}^{(0)}$, $t=0$, $\rho_\mu, \rho_{\mu_{ij}}, \rho_\xi > 1$, $\bar{\mu}$, $\bar{\xi}$, $\mu^{(0)}$, and $\xi^{(0)}$.

**Output:** $\boldsymbol{\mathcal{E}}^{(t+1)}$ and $\boldsymbol{\mathcal{X}}^{(t+1)}$

1. **While not converge do**
2. Calculate $\boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)} (1 \le i < j \le h)$ by
 $\boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)} = \arg\min_{\boldsymbol{\mathcal{Z}}_{i,j}} \mathcal{L}(\boldsymbol{\mathcal{X}}^{(t)}, \boldsymbol{\mathcal{E}}^{(t)}, \boldsymbol{U}_{i,j,k}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}, \boldsymbol{\Lambda}_{i,j}^{(t)}, \boldsymbol{\Lambda}^{(t)}, \mu^{(t)}, \mu_{ij}^{(t)}) + \frac{\xi^{(t)}}{2}\|\boldsymbol{\mathcal{Z}}_{i,j}^{(t)} - \boldsymbol{\mathcal{Z}}_{i,j}\|_F^2$;
3. Update $\boldsymbol{\mathcal{X}}^{(t+1)}$ by
 $\boldsymbol{\mathcal{X}}^{(t+1)} = \arg\min_{\boldsymbol{\mathcal{X}}} \mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}^{(t)}, \boldsymbol{U}_{i,j,k}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)}, \boldsymbol{\Lambda}_{i,j}^{(t)}, \boldsymbol{\Lambda}^{(t)}, \mu^{(t)}, \mu_{ij}^{(t)}) + \frac{\xi^{(t)}}{2}\|\boldsymbol{\mathcal{X}}^{(t)} - \boldsymbol{\mathcal{X}}\|_F^2$;
4. Calculate $\boldsymbol{\mathcal{E}}^{(t+1)}$ by
 $\boldsymbol{\mathcal{E}}^{(t+1)} = \arg\min_{\boldsymbol{\mathcal{E}} \in \mathbb{E}} \mathcal{L}(\boldsymbol{\mathcal{X}}^{(t+1)}, \boldsymbol{\mathcal{E}}, \boldsymbol{U}_{i,j,k}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)}, \boldsymbol{\Lambda}_{i,j}^{(t)}, \boldsymbol{\Lambda}^{(t)}, \mu^{(t)}, \mu_{ij}^{(t)}) + \frac{\xi^{(t)}}{2}\|\boldsymbol{\mathcal{E}}^{(t)} - \boldsymbol{\mathcal{E}}\|_F^2$;
5. Calculate $\boldsymbol{U}_{i,j,k_{n_0}}^{(t+1)}$ $(s+1 \le n \le h, 1 \le i < j \le h)$ by
 $\boldsymbol{U}_{i,j,k_{n_0}}^{(t+1)} = \arg\min_{\boldsymbol{U}_{i,j,k_{n_0}}^T \boldsymbol{U}_{i,j,k_{n_0}} = \boldsymbol{I}} \mathcal{L}(\boldsymbol{\mathcal{X}}^{(t+1)}, \boldsymbol{\mathcal{E}}^{(t+1)}, \{\boldsymbol{U}_{i,j,k_n}\}_{n=s=1}^{n_0}, \boldsymbol{U}_{i,j,k_{n_0}}$
 $, \{\boldsymbol{U}_{i,j,k_n}\}_{n=n_0+1}^h, \boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)}, \boldsymbol{\Lambda}_{i,j}^{(t)}, \boldsymbol{\Lambda}^{(t)}, \mu^{(t)}, \mu_{ij}^{(t)}) + \frac{\xi^{(t)}}{2}\|\boldsymbol{U}_{k_{n_0}}^{(t)} - \boldsymbol{U}_{k_{n_0}}\|_F^2$
6. Calculate $\boldsymbol{\Lambda}_{i,j}^{(t+1)} (1 \le i < j \le h)$ by
 $\boldsymbol{\Lambda}_{i,j}^{(t+1)} = \boldsymbol{\Lambda}_{i,j}^{(t)} + \mu_{ij}^{(t)}(\boldsymbol{\mathcal{Z}}_{i,j} \times_{k_1} \hat{\boldsymbol{U}}_{k_1}^T \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s}^T \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}}^{(t+1)T} \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}^{(t+1)T} - \boldsymbol{\mathcal{X}}^{(t+1)})$
7. Calculate $\boldsymbol{\Lambda}^{(t+1)}$ by $\boldsymbol{\Lambda}^{(t+1)} = \boldsymbol{\Lambda}^{(t)} + \mu^{(t)}(\mathbf{P}_\Omega(\boldsymbol{\mathcal{M}}) - \boldsymbol{\mathcal{X}}^{(t+1)} - \boldsymbol{\mathcal{E}}^{(t+1)})$
8. Update $\mu^{(t+1)}$, $\mu_{ij}^{(t+1)}$ and $\xi^{(t+1)}$ by $\mu^{(t+1)} = \min(\rho_\mu \mu^{(t)}, \bar{\mu})$,
 $\mu_{k_1 k_2}^{(t+1)} = \min(\rho_{k_1 k_2} \mu_{k_1 k_2}^{(t)}, \bar{\mu})$ and $\xi^{(t+1)} = \min(\rho_\xi \xi^{(t)}, \bar{\xi})$, respectively;
9. Check the convergence condition: $\|\boldsymbol{\mathcal{Z}}^{(t+1)} - \boldsymbol{\mathcal{Z}}^{(t)}\|_\infty < \varepsilon$, $\|\boldsymbol{\mathcal{X}}^{(t+1)} - \boldsymbol{\mathcal{X}}^{(t)}\|_\infty < \varepsilon$,
 $\|\boldsymbol{U}_{i,j,k_n}^{(t+1)} - \boldsymbol{U}_{i,j,k_n}^{(t)}\|_\infty < \varepsilon$ for $n = s+1, s+2, \cdots, h$ and $1 \le i < j \le h$;
10. $t = t+1$.
11. **end while**

**Theorem 1.** *For the sequence* $\{[\boldsymbol{\mathcal{X}}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t)}, \{\boldsymbol{U}_{i,j,k_n}^{(t)}\}_{n=s+1}^h, \boldsymbol{\mathcal{E}}^{(t)}, \boldsymbol{\Lambda}^{(t)}, \boldsymbol{\Lambda}_{i,j}^{(t)}, \mu^{(t)}, \mu_{ij}^{(t)}, \xi^{(t)}]\}$ *generated by the proposed algorithm 2, we have the following properties if* $\{\boldsymbol{\Lambda}_{i,j}^{(t)}\}$ *and* $\{\boldsymbol{\Lambda}^{(t)}\}$ *are bounded,* $\sum_{t=1}^\infty (\mu^{(t)})^{-2}\mu^{(t+1)} < +\infty$, $\sum_{t=1}^\infty (\mu_{ij}^{(t)})^{-2}\mu_{ij}^{(t+1)} < +\infty$ *and* $\lim\limits_{n \longrightarrow \infty} \mu^{(n)} \sum\limits_{t=n}^{\infty} (\xi^{(t)})^{-1/2} = \lim\limits_{n \longrightarrow \infty} \mu_{ij}^{(n)} \sum\limits_{t=n}^{\infty} (\xi^{(t)})^{-1/2} = 0$.

*(i)* $\lim\limits_{t \longrightarrow \infty} \mathbf{P}_\Omega(\boldsymbol{\mathcal{M}}) - \boldsymbol{\mathcal{X}}^{(t)} - \boldsymbol{\mathcal{E}}^{(t)} = \mathbf{0}$ *and* $\lim\limits_{t \longrightarrow \infty} \boldsymbol{\mathcal{Z}}_{i,j}^{(t)} - \boldsymbol{\mathcal{X}}^{(t)} \times_{k_1} \hat{\boldsymbol{U}}_{k_1} \cdots \times_{k_s} \hat{\boldsymbol{U}}_{k_s} \times_{k_{s+1}} \boldsymbol{U}_{i,j,k_{s+1}}^{(t)} \cdots \times_{k_h} \boldsymbol{U}_{i,j,k_h}^{(t)} = \mathbf{0}$

*(ii)* $\{[\boldsymbol{\mathcal{X}}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t)}, \{\boldsymbol{U}_{i,j,k_n}^{(t)}\}_{n=s+1}^h\}, \boldsymbol{\mathcal{E}}^{(t)}]\}$ *is bounded.*

*(iii)* $\sum_{t=1}^\infty \xi^{(t)} \|[\boldsymbol{\mathcal{X}}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t)}, \{\boldsymbol{U}_{i,j,k_n}^{(t)}\}_{n=s+1}^h\}, \boldsymbol{\mathcal{E}}^{(t)}] - [\boldsymbol{\mathcal{X}}^{(t+1)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)}, \{\boldsymbol{U}_{i,j,k_n}^{(t+1)}\}_{n=s+1}^h\}, \boldsymbol{\mathcal{E}}^{(t+1)}]\|_F^2$ *is convergent. Thus, we have* $\|[\boldsymbol{\mathcal{X}}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t)}, \{\boldsymbol{U}_{i,j,k_n}^{(t)}\}_{n=s+1}^h\}, \boldsymbol{\mathcal{E}}^{(t)}] - [\boldsymbol{\mathcal{X}}^{(t+1)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t+1)}, \{\boldsymbol{U}_{i,j,k_n}^{(t+1)}\}_{n=s+1}^h\}, \boldsymbol{\mathcal{E}}^{(t+1)}]\|_F^2 \le \mathcal{O}(\frac{1}{\xi^{(t)}})$.

*(iv)* $\lim\limits_{t \longrightarrow \infty} \|\boldsymbol{\Lambda}_{i,j}^{(t+1)} - \boldsymbol{\Lambda}_{i,j}^{(t)}\|_F = 0$ *and* $\lim\limits_{t \longrightarrow \infty} \|\boldsymbol{\Lambda}^{(t+1)} - \boldsymbol{\Lambda}^{(t)}\|_F = 0$.

*(v) Let* $[\boldsymbol{\mathcal{X}}^*, \boldsymbol{\mathcal{Z}}_{i,j}^*, \{\boldsymbol{U}_{i,j,k_n}^*\}_{n=s+1}^h, \boldsymbol{\mathcal{E}}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Lambda}_{i,j}^*]$ *be any limit point of* $\{[\boldsymbol{\mathcal{X}}^{(t)}, \boldsymbol{\mathcal{Z}}_{i,j}^{(t)}, \{\boldsymbol{U}_{i,j,k_n}^{(t)}\}_{n=s+1}^h, \boldsymbol{\mathcal{E}}^{(t)}, \boldsymbol{\Lambda}^{(t)}, \boldsymbol{\Lambda}_{i,j}^{(t)}]\}$. *Then,* $[\boldsymbol{\mathcal{X}}^*, \boldsymbol{\mathcal{Z}}_{i,j}^*, \{\boldsymbol{U}_{i,j,k_n}^*\}_{n=s+1}^h, \boldsymbol{\mathcal{E}}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Lambda}_{i,j}^*]$ *is a KKT point to equation 5.*