Received: 17 June 2019 | Revised: 18 December 2019 | Accepted: 13 February 2020

DOI: 10.1002/nla.2299

RESEARCH ARTICLE

WILEY

Robust tensor completion using transformed tensor singular value decomposition

Guangjing Song[1] | **Michael K. Ng**[2] | **Xiongjun Zhang**[3]

[1]School of Mathematics and Information Sciences, Weifang University, Weifang, China

[2]Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong

[3]School of Mathematics and Statistics and Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan, China

Correspondence
Xiongjun Zhang, School of Mathematics and Statistics and Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan 430079, China.
Email: xjzhang@mail.ccnu.edu.cn

Funding information
Fundamental Research Funds for the Central Universities, Grant/Award Number: CCNU19ZN017; HKRGC GRF, Grant/Award Numbers: 12200317, 12300218, 12300519, 12306616; HKU, Grant/Award Number: 104005583; Hubei Provincial Natural Science Foundation of China, Grant/Award Number: 2018CFB105; National Natural Science Foundation of China, Grant/Award Numbers: 11571098, 11801206, 11871025

Summary

In this article, we study robust tensor completion by using transformed tensor singular value decomposition (SVD), which employs unitary transform matrices instead of discrete Fourier transform matrix that is used in the traditional tensor SVD. The main motivation is that a lower tubal rank tensor can be obtained by using other unitary transform matrices than that by using discrete Fourier transform matrix. This would be more effective for robust tensor completion. Experimental results for hyperspectral, video and face datasets have shown that the recovery performance for the robust tensor completion problem by using transformed tensor SVD is better in peak signal-to-noise ratio than that by using Fourier transform and other robust tensor completion methods.

KEYWORDS

low-rank, robust tensor completion, sparsity, transformed tensor singular value decomposition, unitary transform matrix

MOS SUBJECT CLASSIFICATION

15A04; 65F99; 90C25

1 | INTRODUCTION

Tensors (multidimensional arrays) are generalizations of vectors and matrices, which can be used as a powerful tool in modeling multidimensional data such as videos,[1] color images,[2,3] hyperspectral images,[4–6] and electroencephalography.[7] Based on its multilinear algebraic properties, a tensor can take full advantage of its structures to provide better understanding and higher accuracy of the multidimensional data. In many real-world applications, tensor datasets are often corrupted and/or incomplete owing to various unpredictable or unavoidable situations.[3,8–15] It motivated us to perform tensor completion and robust tensor completion for multidimensional data processing.

As a special case of tensor completion, the matrix completion problem, which aims to recover a low rank matrix from a subset of sampled entries, has received much attention. Note that the nuclear norm of a matrix is the convex envelope of the rank over a unit ball of spectral norm,[16,17] then one can relax the nonconvex rank minimization problem to a convex

nuclear norm minimization problem. In the seminal articles,[18,19] by solving the relaxed convex programming problems, the authors have recovered a low rank matrix exactly with overwhelming probability, from a small fraction of its entries, even part of them are corrupted, provided that the corruptions are reasonably sparse. Some other results can be found in References 20–22 and references therein.

Compared with the rapid growth of the applications of real-world multidimensional data, the tensor completion problem is far from being well-studied. Intuitively, the tensor completion problem could be solved by unfolding a tensor into a large-scale matrix. However, it has been shown that the matrix completion methods may break the multiway structure of a tensor and lose the redundancy among modes. Therefore, it is important and meaningful to consider the tensor completion problem without unfolding it into a matrix directly. Being different from the matrix case, there are many kinds of tensor rank definitions which lead to different low rank tensor completion models. For instance, the CANDECOMP/PARAFAC (CP) rank is defined as the minimal number of the rank one outer products of tensors,[23] which is NP-hard to compute in general.[24] Although many authors[25,26] have recovered some special low CP rank tensors by different methods, it is often computationally intractable to determine the CP rank or its best convex approximation. The tensor train (TT) rank[27] is generated by the TT decomposition using the link structure of each core tensor. Since the link structure, the TT rank is only efficient for higher order tensor for tensor completion. Bengua et al.[28] proposed a novel approach based on TT rank for color images and videos completion. However, this method may be challenged when the third-dimension of the data is high, such as hyperspectral data. The Tucker rank (multirank) is actually a vector whose entries can be derived from the factors of Tucker decomposition.[29] Liu et al.[1] proposed to use the sum of the nuclear norms of unfolding matrices of a tensor to recover a low Tucker rank tensor. However, the sum of the nuclear norms is not the convex envelope of the sum of ranks of those unfolding matrices.[30] Mu et al.[31] showed that the method discussed in Reference 1 is suboptimal and proposed a square deal method to recover a low rank and high-order tensor. While the square deal method only utilizes one mode information of unfolding matrices for third-order tensors. Other extensions can be found in Reference 32 and references therein. In Reference 33, Gu et al. provided a perfect recovery of two components (the low-rank tensor and the entrywise sparse corruption tensor) under restricted eigenvalue conditions. In Reference 34, Huang et al. proposed a tensor robust principal component analysis model for exact recovery guarantee under certain tensor incoherence conditions.

The tensor-tensor product (t-product) and associated algebraic construction based on the Fourier transform, cosine transform, and any invertible transform for tensors of order three or higher are studied in References 35–37, respectively. With this framework, Kilmer et al.[35] introduced a factorization for third-order tensors called the tensor singular value decomposition (SVD) as well as the tensor tubal rank. Compared with other tensor decompositions, this tensor SVD has been shown to be superior in capturing the spatial-shifting correlation that is ubiquitous in real-world data.[35,37–40] Moreover, the tubal nuclear norm is the convex envelope of the tubal average rank within the unit ball of the tensor spectral norm.[39] Motivated by the above results, Zhang et al.[41] derived theoretical performance bounds of the model proposed in Reference 39 using the tensor SVD algebraic framework for third-order tensor recovery from limited sampling. Zhou et al.[42] proposed a novel factorization method based on the tensor nuclear norm in the Fourier domain for solving the third-order tensor completion problem. Hu et al.[43] proposed a twist tensor nuclear norm for tensor completion, which relaxes the tensor multirank of the twist tensor in the Fourier domain. Being different from tensor completion, robust tensor completion is more complex due to the sparse noise in the observations. Jiang et al.[44] showed that one can recover a low tubal rank tensor exactly with overwhelming probability by simply solving a convex program, where the objective function is a weighted combination of the tubal nuclear norm, a convex surrogate of the tubal rank, and the ℓ_1-norm. Recently, Lu et al.[45] considered the tensor robust principal component analysis problem and proposed a tensor nuclear norm based on t-product and tensor SVD in the Fourier domain, where the theoretical guarantee for exact recovery was also provided.

The main aim of this article is to study the robust tensor completion problem by using transformed tensor SVD, which employs unitary transform matrices instead of discrete Fourier transform matrix in the tensor SVD. The transformation can redistribute tubal entries into the matrix slices (along the tube direction) of tensors, and then SVD is applied to each transformed matrix slice to find out their decomposition. The main motivation is that with suitable transformation, low rank transformed matrix slices can be resulted. Therefore, a lower tubal rank tensor can be obtained by using other unitary transform matrices than that by using discrete Fourier transform matrix. For example, in Figures 1, 3, and 5 of our experimental results in Section 4, we observe that there are more small numbers in singular tubes by using other unitary transform matrices that those by using discrete Fourier transform matrix. This would be more effective for robust tensor completion as a low rank requirement is preferred (see Theorem 2).

Here we summarize our contributions of this article as follows. (i) We generalize low tubal rank recovery theoretical results for any unitary transformation. More precisely, we can recover a low transformed tubal rank tensor

exactly with overwhelming probability provided that its tubal rank is sufficiently small and its corrupted entries are reasonably sparse. Because of the use of unitary transformation, there are new results related to the convex envelope of the tensor rank, the subgradient formula and tensor basis required in the proof. (ii) We propose a new unitary transformation that can lead to significant recovery results compared with the use of the Fourier transform. (iii) Experimental results for hyperspectral and face images and video data have shown that the recovery performance by using transformed tensor SVD is better in peak signal-to-noise ratio (PSNR) than that by using Fourier transform in tensor SVD and other tensor completion methods.

The outline of this article is given as follows. In Section 2, we introduce transformed tensor SVD. In Section 3, we analyze the robust tensor completion problem and the algorithm for solving the model. In Section 4, numerical results are presented to show the effectiveness of the proposed tensor SVD for the robust tensor completion problem. Finally, some concluding remarks are given in Section 5. All proofs are deferred to the Appendix.

1.1 | Notation and preliminaries

Throughout this article, the fields of real number and complex number are denoted as $\mathbb{R}$ and $\mathbb{C}$, respectively. Tensors and matrices are denoted by Euler letters and boldface capital letters, respectively. For a third-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote its (i,j,k)th entry as $\mathcal{A}_{ijk}$ and use the MATLAB notation $\mathcal{A}(i,:,:), \mathcal{A}(:,i,:)$ and $\mathcal{A}(:,:,i)$ to denote the ith horizontal, lateral, and frontal slices, respectively. Specifically, the frontal slice $\mathcal{A}(:,:,i)$ is denoted compactly as $\mathcal{A}^{(i)}$. $\mathcal{A}(i,j,:)$ denotes a tubal fiber obtained by fixing the first two indices and varying the third index. Moreover, a tensor tube of size $1 \times 1 \times n_3$ is denoted as $\mathring{\boldsymbol{a}}$ and a tensor column of size $n_1 \times 1 \times n_3$ is denoted as $\vec{a}$.

The inner product of $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n_1 \times n_2}$ is given by $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{Tr}(\mathbf{A}^H \mathbf{B})$, where $\mathbf{A}^H$ denotes the conjugate transpose of $\mathbf{A}$ and $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. For a vector $v \in \mathbb{C}^n$, the ℓ_2-norm is $\|v\|_2 = \sqrt{\sum_i |v_i|^2}$. The spectral norm of a matrix $\mathbf{A} \in \mathbb{C}^{n_1 \times n_2}$ is denoted as $\|\mathbf{A}\| = \max_i \sigma_i(\mathbf{A})$, where $\sigma_i(\mathbf{A})$ is the ith largest singular value of $\mathbf{A}$. The nuclear norm of a matrix is defined as $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$. For a tensor $\mathcal{A}$, the ℓ_1-norm is defined as $\|\mathcal{A}\|_1 = \sum_{i,j,k} |\mathcal{A}_{ijk}|$, the infinity norm is defined as $\|\mathcal{A}\|_\infty = \max_{i,j,k} |\mathcal{A}_{ijk}|$ and the Frobenius norm is defined as $\|\mathcal{A}\|_F = \sqrt{\sum_{i,j,k} |\mathcal{A}_{ijk}|^2}$. Suppose that $\mathtt{L}$ is a tensor operator, then its operator norm is defined as $\|\mathtt{L}\|_{\mathrm{op}} = \sup_{\|\mathcal{A}\|_F \le 1} \|\mathtt{L}(\mathcal{A})\|_F$.

2 | TRANSFORMED TENSOR SVD

Let $\mathbf{\Phi} \in \mathbb{C}^{n_3 \times n_3}$ be a unitary transform matrix with $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}_{n_3}$, where $\mathbf{I}_{n_3}$ is the $n_3 \times n_3$ identity matrix. $\hat{\mathcal{A}}_{\mathbf{\Phi}}$ represents a third-order tensor obtained via multiplying by $\mathbf{\Phi}$ on all tubes along the third-dimension of $\mathcal{A}$, that is,

$$\hat{\mathcal{A}}_{\mathbf{\Phi}}(i,j,:) = \mathbf{\Phi}(\mathcal{A}(i,j,:)).$$

Here we write $\hat{\mathcal{A}}_{\mathbf{\Phi}} = \mathbf{\Phi}[\mathcal{A}]$. Moreover, one can get $\mathcal{A}$ from $\hat{\mathcal{A}}_{\mathbf{\Phi}}$ by using $\mathbf{\Phi}^H$ operation along the third-dimension of $\hat{\mathcal{A}}_{\mathbf{\Phi}}$, that is, $\mathcal{A} = \mathbf{\Phi}^H[\hat{\mathcal{A}}_{\mathbf{\Phi}}]$.

We construct a block diagonal matrix based on the frontal slices of $\hat{\mathcal{A}}_{\mathbf{\Phi}}$ as follows:

$$\overline{\mathcal{A}_{\mathbf{\Phi}}} = \mathrm{blockdiag}(\hat{\mathcal{A}}_{\mathbf{\Phi}}) := \begin{pmatrix} \hat{\mathcal{A}}_{\mathbf{\Phi}}^{(1)} & & & \\ & \hat{\mathcal{A}}_{\mathbf{\Phi}}^{(2)} & & \\ & & \ddots & \\ & & & \hat{\mathcal{A}}_{\mathbf{\Phi}}^{(n_3)} \end{pmatrix},$$

The block diagonal matrix can be converted into a tensor by the following fold operator:

$$\mathrm{fold}(\mathrm{blockdiag}(\hat{\mathcal{A}}_{\mathbf{\Phi}})) = \hat{\mathcal{A}}_{\mathbf{\Phi}}.$$

Kernfeld et al.[36] defined the $\star_{\mathbf{L}}$-product between two tensors by the slices products in the transformed domain, where $\mathbf{L}$ is an arbitrary invertible transform. In this article, we are mainly interested in the t-product which is based on unitary transformations.

Definition 1. The $\mathbf{\Phi}$-product of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{C}^{n_2 \times n_4 \times n_3}$ is a tensor $\mathcal{C} \in \mathbb{C}^{n_1 \times n_4 \times n_3}$, which is given by

$$\mathcal{C} = \mathcal{A} \diamond_{\mathbf{\Phi}} \mathcal{B} = \mathbf{\Phi}^H \left[\text{fold} \left(\text{blockdiag}(\hat{\mathcal{A}}_{\mathbf{\Phi}}) \times \text{blockdiag}(\hat{\mathcal{B}}_{\mathbf{\Phi}}) \right) \right],$$

where "$\times$" denotes the standard matrix product.

The t-product[35] of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is a tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ given by

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \text{Fold}_{\text{vec}} \left(\text{Circ}(\mathcal{A}) \times \text{Vec}(\mathcal{B}) \right), \tag{1}$$

where Fold_{vec} is an operation that takes $\text{Vec}(\mathcal{B})$ into a tensor, that is, $\text{Fold}_{\text{vec}}(\text{Vec}(\mathcal{B})) = \mathcal{B}$,

$$\text{Vec}(\mathcal{B}) = \begin{pmatrix} \mathcal{B}^{(1)} \\ \mathcal{B}^{(2)} \\ \vdots \\ \mathcal{B}^{(n_3)} \end{pmatrix},$$

and

$$\text{Circ}(\mathcal{A}) = \begin{pmatrix} \mathcal{A}^{(1)} & \mathcal{A}^{(n_3)} & \mathcal{A}^{(n_3-1)} & \cdots & \mathcal{A}^{(2)} \\ \mathcal{A}^{(2)} & \mathcal{A}^{(1)} & \mathcal{A}^{(n_3)} & \cdots & \mathcal{A}^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{A}^{(n_3)} & \mathcal{A}^{(n_3-1)} & \cdots & \mathcal{A}^{(2)} & \mathcal{A}^{(1)} \end{pmatrix}.$$

The t-product (1) can be seen as a special case of Definition 1. Recall that the block circulant matrix $\text{Circ}(\mathcal{A})$ can be diagonalized by the discrete Fourier transform matrix $\mathbf{F}_{n_3}$ and the block diagonal matrices are the frontal slices of $\hat{\mathcal{A}}_{\mathbf{F}_{n_3}}$, that is,

$$(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) \times \text{Circ}(\mathcal{A}) \times (\mathbf{F}_{n_3}^H \otimes \mathbf{I}_{n_2}) = \text{blockdiag}(\hat{\mathcal{A}}_{\mathbf{F}_{n_3}}),$$

where $\otimes$ is the Kronecker product. It follows that

$$\begin{aligned}
\mathcal{A} * \mathcal{B} &= \text{Fold}_{\text{vec}} \left(\text{Circ}(\mathcal{A}) \times \text{Vec}(\mathcal{B}) \right) \\
&= \text{Fold}_{\text{vec}} \left(\left(\mathbf{F}_{n_3}^H \otimes \mathbf{I}_{n_1} \right) \times \text{blockdiag} \left(\hat{\mathcal{A}}_{\mathbf{F}_{n_3}} \right) \times \left(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_2} \right) \times \text{Vec}(\mathcal{B}) \right) \\
&= \text{Fold}_{\text{vec}} \left(\left(\mathbf{F}_{n_3}^H \otimes \mathbf{I}_{n_1} \right) \times \text{blockdiag} \left(\hat{\mathcal{A}}_{\mathbf{F}_{n_3}} \right) \times \text{Vec} \left(\hat{\mathcal{B}}_{\mathbf{F}_{n_3}} \right) \right) \\
&= \text{fold} \left(\left(\mathbf{F}_{n_3}^H \otimes \mathbf{I}_{n_1} \right) \times \text{blockdiag} \left(\hat{\mathcal{A}}_{\mathbf{F}_{n_3}} \right) \times \text{blockdiag} \left(\hat{\mathcal{B}}_{\mathbf{F}_{n_3}} \right) \right) \\
&= \mathbf{F}_{n_3}^H \left[\text{fold} \left(\text{blockdiag} \left(\hat{\mathcal{A}}_{F_{n_3}} \right) \times \text{blockdiag} \left(\hat{\mathcal{B}}_{\mathbf{F}_{n_3}} \right) \right) \right] \\
&= \mathcal{A} \diamond_{\mathbf{F}_{n_3}} \mathcal{B}.
\end{aligned}$$

The definitions of the conjugate transpose of $\mathcal{A}$, the identity tensor, the unitary tensor, and the inner product between two tensors related to the $\mathbf{\Phi}$-product, can be given as follows.

Definition 2. The conjugate transpose of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with respect to $\mathbf{\Phi}$ is the tensor $\mathcal{A}^H \in \mathbb{C}^{n_2 \times n_1 \times n_3}$ obtained by

$$\mathcal{A}^H = \mathbf{\Phi}^H \left[\text{fold} \left(\text{blockdiag}(\hat{\mathcal{A}}_{\mathbf{\Phi}})^H \right) \right].$$

Definition 3 (Proposition 4.1 in Reference 36). The identity tensor $\mathcal{I}_{\mathbf{\Phi}} \in \mathbb{C}^{n \times n \times n_3}$ (with respect to $\mathbf{\Phi}$) is defined to be a tensor such that $\mathcal{I}_{\mathbf{\Phi}} = \mathbf{\Phi}^H[\mathcal{T}]$, where $\mathcal{T} \in \mathbb{R}^{n \times n \times n_3}$ with each frontal slice being the $n \times n$ identity matrix.

Definition 4 (Definition 5.1 in Reference 36). A tensor $\mathcal{Q} \in \mathbb{C}^{n \times n \times n_3}$ is unitary with respect to $\mathbf{\Phi}$-product if it satisfies

$$\mathcal{Q}^H \diamond_{\mathbf{\Phi}} \mathcal{Q} = \mathcal{Q} \diamond_{\mathbf{\Phi}} \mathcal{Q}^H = \mathcal{I}_{\mathbf{\Phi}},$$

where $\mathcal{I}_{\mathbf{\Phi}}$ is the identity tensor.

Definition 5. The inner product of $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{n_3} \langle \mathcal{A}^{(i)}, \mathcal{B}^{(i)} \rangle = \langle \overline{\mathcal{A}_{\mathbf{\Phi}}}, \overline{\mathcal{B}_{\mathbf{\Phi}}} \rangle, \tag{2}$$

where $\langle \mathcal{A}^{(i)}, \mathcal{B}^{(i)} \rangle$ is the standard inner product of two matrices.

In addition, a tensor $\mathcal{A}$ is said to be diagonal if each frontal slice $\mathcal{A}^{(i)}$ is a diagonal matrix.[35] Based on the above definitions, we have the following transformed tensor SVD with respect to $\mathbf{\Phi}$.

Theorem 1 (Theorem 5.1 in Reference 36). *Suppose that $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$. Then $\mathcal{A}$ can be factorized as follows:*

$$\mathcal{A} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \tag{3}$$

where $\mathcal{U} \in \mathbb{C}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{C}^{n_2 \times n_2 \times n_3}$ are unitary tensors with respect to $\mathbf{\Phi}$-product, and $\mathcal{S} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is a diagonal tensor.

The tensors $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{S}$ in the transformed tensor SVD can be computed by SVDs of $\hat{\mathcal{A}}_{\mathbf{\Phi}}^{(i)}$, which is summarized in Algorithm 1.

Algorithm 1. Transformed tensor SVD for third-order tensors[36]

Input: $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$.
1: $\hat{\mathcal{A}}_{\mathbf{\Phi}} = \mathbf{\Phi}[\mathcal{A}]$;
2: **for** $i = 1, ..., n_3$ **do**
3: $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \mathrm{SVD}(\hat{\mathcal{A}}_{\mathbf{\Phi}}^{(i)})$;
4: $\hat{\mathcal{U}}_{\mathbf{\Phi}}^{(i)} = \mathbf{U}, \hat{\mathcal{S}}_{\mathbf{\Phi}}^{(i)} = \mathbf{S}, \hat{\mathcal{V}}_{\mathbf{\Phi}}^{(i)} = \mathbf{V}$;
5: **end for**
6: $\mathcal{U} = \mathbf{\Phi}^H[\hat{\mathcal{U}}_{\mathbf{\Phi}}]$, $\mathcal{S} = \mathbf{\Phi}^H[\hat{\mathcal{S}}_{\mathbf{\Phi}}]$, $\mathcal{V} = \mathbf{\Phi}^H[\hat{\mathcal{V}}_{\mathbf{\Phi}}]$.
Output: $\mathcal{U} \in \mathbb{C}^{n_1 \times n_1 \times n_3}$, $\mathcal{S} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, $\mathcal{V} \in \mathbb{C}^{n_2 \times n_2 \times n_3}$.

Remark 1. For computational improvement, we also use the skinny transformed tensor SVD. For any $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with $\mathrm{rank}(\mathcal{A}) = r$ (see in the following definition), the skinny transformed tensor SVD is given by $\mathcal{A} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H$, where $\mathcal{U} \in \mathbb{C}^{n_1 \times r \times n_3}$, $\mathcal{V} \in \mathbb{C}^{n_2 \times r \times n_3}$ satisfying $\mathcal{U}^H \diamond_{\mathbf{\Phi}} \mathcal{U} = \mathcal{I}_{\mathbf{\Phi}}$, $\mathcal{V}^H \diamond_{\mathbf{\Phi}} \mathcal{V} = \mathcal{I}_{\mathbf{\Phi}}$, and $\mathcal{S} \in \mathbb{C}^{r \times r \times n_3}$ is a diagonal tensor.

Based on the transformed tensor SVD given in Theorem 1, the transformed multirank and tubal rank of a tensor can be defined as follows.

Definition 6. The transformed multirank of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with respect to $\mathbf{\Phi}$ is a vector $\boldsymbol{r} \in \mathbb{R}^{n_3}$ with its ith entry being the rank of the ith frontal slice of $\hat{\mathcal{A}}_{\mathbf{\Phi}}$, that is, $r_i = \mathrm{rank}(\hat{\mathcal{A}}_{\mathbf{\Phi}}^{(i)})$. The transformed tubal rank, denoted as $\mathrm{rank}(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ comes from the transformed tensor SVD of $\mathcal{A} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H$, that is,

$$\mathrm{rank}(\mathcal{A}) = \#\{i : \mathcal{S}(i, i, :) \neq \mathbf{0}\} = \max_i r_i, \tag{4}$$

where # denotes the cardinality of a set.

It follows from Reference 44 that the tensor spectral norm of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with respect to $\mathbf{\Phi}$, denoted as $\|\mathcal{A}\|_{\mathbf{\Phi}}$, can be defined as $\|\mathcal{A}\|_{\mathbf{\Phi}} = \|\overline{\mathcal{A}_{\mathbf{\Phi}}}\|$. In other words, the tensor spectral norm of $\mathcal{A}$ equals to the matrix spectral norm of its block diagonal form in the transformed domain. Moreover, suppose that a tensor operator $\mathtt{L}$ can be represented as a tensor $\mathcal{L}$ via Φ-product, we have $\|\mathtt{L}\|_{\mathrm{op}} = \|\mathcal{L}\|_{\mathbf{\Phi}}$. The aim of this article is to recover a low transformed tubal rank tensor, which motivates us to introduce the following definition, that is, the transformed tubal nuclear norm (TTNN) of a tensor.

Definition 7. The TTNN of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{A}\|_{\mathrm{TTNN}}$, is the sum of the nuclear norms of all frontal slices of $\hat{\mathcal{A}}_{\mathbf{\Phi}}$, that is, $\|\mathcal{A}\|_{\mathrm{TTNN}} = \sum_{i=1}^{n_3} \|\hat{\mathcal{A}}_{\mathbf{\Phi}}^{(i)}\|_*$.

The following lemma shows that the TTNN of a tensor is the convex envelope of the sum of the entries of the transformed multirank over a unit ball of the tensor spectral norm. This is why the TTNN is useful for low transformed tubal rank tensor recovery. We remark this is the new result in the literature, and the proof is different from theorem 1 of Reference 16 because we consider the complex-valued matrices and tensors.

Lemma 1. *For any tensor* $\mathcal{X} \in \mathbb{C}^{n_1\times n_2\times n_3}$, $\text{rank}_{sum}(\mathcal{X}) = \sum_{i=1}^{n_3} \text{rank}(\hat{\mathcal{X}}_{\mathbf{\Phi}}^{(i)})$ *denotes a transformed multirank function. Then* $\|\mathcal{X}\|_{\text{TTNN}}$ *is the convex envelope of the function* $\text{rank}_{sum}(\mathcal{X})$ *on the set* $\{\mathcal{X} \mid \|\mathcal{X}\|_{\mathbf{\Phi}} \le 1\}$.

The proof can be found in Appendix A. Next we will introduce two kinds of tensor bases which will be used in introducing the tensor coordinate decomposition and tensor incoherence conditions.

Definition 8. (i) The transformed column basis with respect to $\mathbf{\Phi}$, which is a lateral slice and denoted as $\vec{\boldsymbol{e}}_{ik}$, is a matrix of size $n_1 \times n_3$ with the (i,k)th element equaling to 1 and the others equaling to 0. (ii) Denote $\mathring{\boldsymbol{e}}_k$, which is a tube, as a vector of size n_3 with the kth entry equaling to 1 and the remaining entries equaling to 0, $\mathbf{\Phi}(\mathring{\boldsymbol{e}}_k)_t$ as the tth entry of $\mathbf{\Phi}(\mathring{\boldsymbol{e}}_k)$, $t = 1, \dots, n_3$. The transformed tube basis with respect to $\mathbf{\Phi}$, which is a tube and denoted as $\mathbf{\Phi}(\mathring{\boldsymbol{e}}_{kk})$, is a vector of size n_3 with the tth entry of $\mathbf{\Phi}(\mathring{\boldsymbol{e}}_{kk})$ equaling to $(\mathbf{\Phi}(\mathring{\boldsymbol{e}}_k)_t)^{-1}$, if $\mathbf{\Phi}(\mathring{\boldsymbol{e}}_k)_t \neq 0$, and 0 otherwise, $t = 1, \dots, n_3$.

The new bases are different from existing bases, and are useful to the proof of tensor recovery via transformed tensor SVD. Let $\mathcal{E}_{ijk}$ be a unit tensor whose (i,j,k)th entry is 1 and others are 0. Recall Definition 8, $\mathcal{E}_{ijk}$ can be expressed as $\mathcal{E}_{ijk} = \vec{\boldsymbol{e}}_{ik} \diamond_{\mathbf{\Phi}} \mathring{\boldsymbol{e}}_{kk} \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{jk}^{H}$, then for an arbitrary tensor $\mathcal{Z} \in \mathbb{C}^{n_1\times n_2\times n_3}$, we have

$$\mathcal{Z} = \sum_{ijk} \langle \mathcal{E}_{ijk}, \mathcal{Z} \rangle \mathcal{E}_{ijk} = \sum_{ijk} \left\langle \vec{\boldsymbol{e}}_{ik} \diamond_{\mathbf{\Phi}} \mathring{\boldsymbol{e}}_{kk} \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{jk}^{H}, \mathcal{Z} \right\rangle \vec{\boldsymbol{e}}_{ik} \diamond_{\mathbf{\Phi}} \mathring{\boldsymbol{e}}_{kk} \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{jk}^{H}.$$

This property will be used many times in the proof of our main results in Section 3.

3 | RECOVERY RESULTS BY TRANSFORMED TENSOR SVD

Suppose that we are given a third-order tensor $\mathcal{L}_0$ that has low transformed tubal rank with respect to $\mathbf{\Phi}$ and is also corrupted by a sparse tensor $\mathcal{E}_0$, where the transformed tubal rank of $\mathcal{L}_0$ is unknown. Moreover, we have no idea about the locations of the nonzero entries of $\mathcal{E}_0$, not even how many there are. Our aim is to recover $\mathcal{L}_0$ from a set of the observed entries of $\mathcal{X}$. The TTNN of a tensor is used to get a low rank solution and the ℓ_1-norm is used to get a sparse solution. Mathematically, the model can be stated as follows:

$$\min_{\mathcal{L},\,\mathcal{E}} \ \|\mathcal{L}\|_{\text{TTNN}} + \lambda\|\mathcal{E}\|_1, \quad \text{s.t.,}\ \mathcal{P}_\Omega(\mathcal{L}+\mathcal{E}) = \mathcal{P}_\Omega(\mathcal{X}), \tag{5}$$

where λ is a penalty parameter and $\mathcal{P}_\Omega$ is a linear projection such that the entries in the set Ω are given while the remaining entries are missing.

We remark that the convex optimization problems constructed in References 39,41, and 44 can be seen as special cases of (5), which aim to solve the tensor completion and tensor robust principal component analysis, respectively. For instance, if the unitary transform $\mathbf{\Phi}$ is chosen as the discrete Fourier transform, the TTNN can be replaced by the tubal nuclear norm.

Here we need some transformed tensor incoherence conditions on $\mathcal{L}_0$ to ensure that it is not sparse.

Definition 9. Assume that $\text{rank}(\mathcal{L}_0) = r$ and its skinny transformed tensor SVD is $\mathcal{L}_0 = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H \in \mathbb{C}^{n_1\times n_2\times n_3}$. $\mathcal{L}_0$ is said to satisfy the transformed tensor incoherence conditions with parameter $\mu > 1$ if

$$\max_{i=1,\dots,n_1} \max_{k=1,\dots,n_3} \|\mathcal{U}^H \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{ik}\|_F \le \sqrt{\frac{\mu r}{n_1}}, \tag{6}$$

$$\max_{j=1,\dots,n_2} \max_{k=1,\dots,n_3} \|\mathcal{V}^H \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{jk}\|_F \le \sqrt{\frac{\mu r}{n_2}}, \tag{7}$$

and

$$\|\mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H\|_\infty \le \sqrt{\frac{\mu r}{n_1 n_2 n_3}}, \tag{8}$$

where $\vec{\boldsymbol{e}}_{ik}$ and $\vec{\boldsymbol{e}}_{jk}$ are the tensor basis with respect to the transformation $\mathbf{\Phi}$.

For convenience, we denote $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$. The main result of this article can be stated in the following theorem.

Theorem 2. *Suppose that $\mathcal{L}_0 \in \mathbb{C}^{n_1\times n_2\times n_3}$ obeys (6)–(8), and the observation set Ω is uniformly distributed among all sets of cardinality $m = \rho n_1 n_2 n_3$. Also suppose that each observed entry is independently corrupted with probability γ. Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1(n_{(1)}n_3)^{-c_2}$, the recovery of $\mathcal{L}_0$ with $\lambda = 1/\sqrt{\rho n_{(1)} n_3}$ is exact, provided that*

$$r \le \frac{c_r n_{(2)}}{\mu(\log(n_{(1)}n_3))^2} \quad \text{and} \quad \gamma \le c_\gamma, \tag{9}$$

where c_r and c_γ are two positive constants.

Remark 2. By the inner product given in Definition 5, a direct generalization of the transformed tensor incoherence conditions listed in Reference 44 for arbitrary unitary transformations are

$$\max_{i=1,\dots,n_1} \|\mathcal{U}^H \diamond_{\mathbf{\Phi}} \vec{\mathbf{e}}_{i1}\|_F \le \sqrt{\frac{\mu r}{n_1 n_3}}, \quad \max_{j=1,\dots,n_2} \|\mathcal{V}^H \diamond_{\mathbf{\Phi}} \vec{\mathbf{e}}_{j1}\|_F \le \sqrt{\frac{\mu r}{n_2 n_3}},$$

and

$$\|\mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H\|_\infty \le \sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}.$$

The right hands of the three inequalities are obviously smaller than those given in Equations (6)–(8), which means that the exact recovery conditions given in Theorem 2 are weaker than those given in Reference 44.

The idea of the proof is to employ convex analysis to derive the conditions in which one can check whether the pair $(\mathcal{L}, \mathcal{E})$ is the unique minimizer to (5), and to explicitly show that the conditions in Theorem 2 are met with overwhelming probability. The main tools of our proof are the noncommutative Bernstein Inequality and the golfing scheme.[18,46] The detailed proof is given in Appendix B.

3.1 | Optimization algorithm

In this subsection, we develop a symmetric Gauss-Seidel-based multiblock alternating direction method of multipliers (sGS-ADMM)[47,48] to solve the robust tensor completion problem (5). The sGS-ADMM has validated the efficiency in many fields, for example, see References 40,47–50 and references therein. Let $\mathcal{L} + \mathcal{E} = \mathcal{M}$. Problem (5) can be rewritten as

$$\begin{aligned} \min_{\mathcal{L},\ \mathcal{E},\ \mathcal{M}} \ & \|\mathcal{L}\|_{\mathrm{TTNN}} + \lambda\|\mathcal{E}\|_1 \\ \text{s.t. } & \mathcal{L} + \mathcal{E} = \mathcal{M},\ \mathcal{P}_\Omega(\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}). \end{aligned} \tag{10}$$

The augmented Lagrangian function of (10) is defined by

$$L(\mathcal{L},\ \mathcal{E},\ \mathcal{M},\ \mathcal{Z}) := \|\mathcal{L}\|_{\mathrm{TTNN}} + \lambda\|\mathcal{E}\|_1 - \langle \mathcal{Z}, \mathcal{L} + \mathcal{E} - \mathcal{M}\rangle + \frac{\beta}{2}\|\mathcal{L} + \mathcal{E} - \mathcal{M}\|_F^2,$$

where $\mathcal{Z}$ is the Lagrangian multiplier and $\beta > 0$ is the penalty parameter. Let $\mathfrak{D} := \{\mathcal{M} \in \mathbb{C}^{n_1\times n_2\times n_3} : \mathcal{P}_\Omega(\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X})\}$. The iteration system of the sGS-ADMM can be described as follows:

$$\mathcal{M}^{k+\frac{1}{2}} = \arg\min\nolimits_{\mathcal{M}\in\mathfrak{D}} \left\{L(\mathcal{L}^k,\ \mathcal{E}^k,\ \mathcal{M},\ \mathcal{Z}^k)\right\}, \tag{11}$$

$$\mathcal{L}^{k+1} = \arg\min_{\mathcal{L}} \left\{L(\mathcal{L},\ \mathcal{E}^k,\ \mathcal{M}^{k+\frac{1}{2}},\ \mathcal{Z}^k)\right\}, \tag{12}$$

$$\mathcal{M}^{k+1} = \arg\min\nolimits_{\mathcal{M}\in\mathfrak{D}} \left\{L(\mathcal{L}^{k+1},\ \mathcal{E}^k,\ \mathcal{M},\ \mathcal{Z}^k)\right\}, \tag{13}$$

$$\mathcal{E}^{k+1} = \arg\min_{\mathcal{E}} \left\{L(\mathcal{L}^{k+1},\ \mathcal{E},\ \mathcal{M}^{k+1},\ \mathcal{Z}^k)\right\}, \tag{14}$$

$$\mathcal{Z}^{k+1} = \mathcal{Z}^k - \tau\beta(\mathcal{L}^{k+1} + \mathcal{E}^{k+1} - \mathcal{M}^{k+1}), \tag{15}$$

where $\tau \in (0, (1+\sqrt{5})/2)$ is the step-length.

The solution with respect to $\mathcal{M}$ can be given by

$$\mathcal{M} = \begin{cases} \mathcal{X}_{ijk}, & \text{if } (i,j,k) \in \Omega, \\ \left(\mathcal{L} + \mathcal{E} - \frac{1}{\beta}\mathcal{Z}\right)_{ijk}, & \text{otherwise.} \end{cases} \tag{16}$$

Similar to the proximal mapping of the nuclear norm of a matrix, we give the proximal mapping of the TTNN of a tensor. The proximal mapping of $\|\cdot\|_{\text{TTNN}}$ at $\mathcal{Y}$ can be given in the following theorem.

Theorem 3. *For any $\mathcal{Y} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the minimizer of the following problem*

$$\min_{\mathcal{X}} \left\{ \lambda \|\mathcal{X}\|_{\text{TTNN}} + \frac{1}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2 \right\}, \tag{17}$$

is given by

$$\text{Prox}_{\lambda\|\cdot\|_{\text{TTNN}}}(\mathcal{Y}) := \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S}_\lambda \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \tag{18}$$

where $\mathcal{S}_\lambda = \mathbf{\Phi}^H[\hat{\mathcal{S}}_\lambda]$ and $\hat{\mathcal{S}}_\lambda = \max\{\hat{\mathcal{S}}_{\mathbf{\Phi}} - \lambda, 0\}$.

By the definition of the TTNN, problem (17) can be rewritten as

$$\min_{\mathcal{X}} \left\{ \sum_{i=1}^{n_3} \lambda \|\hat{\mathcal{X}}_{\mathbf{\Phi}}^{(i)}\|_* + \frac{1}{2}\|\hat{\mathcal{X}}_{\mathbf{\Phi}}^{(i)} - \hat{\mathcal{Y}}_{\mathbf{\Phi}}^{(i)}\|_F^2 \right\}. \tag{19}$$

By theorem 2.1 in Reference 51, the minimizer of (19) is given by

$$\hat{\mathcal{X}}_{\mathbf{\Phi}}^{(i)} = \hat{\mathcal{U}}_{\mathbf{\Phi}}^{(i)} \hat{\mathbf{\Sigma}}_\lambda (\hat{\mathcal{V}}_{\mathbf{\Phi}}^{(i)})^H,$$

where $\hat{\mathbf{\Sigma}}_\lambda = \max\{\hat{\mathcal{S}}_{\mathbf{\Phi}}^{(i)} - \lambda, 0\}$. By using the inverse unitary transform along the third-dimension, we get that the optimal solution of (17) is given by (18).

Note that the subproblem with respect to $\mathcal{L}$ in (12) can be described as

$$\min \|\mathcal{L}\|_{\text{TTNN}} + \frac{\beta}{2}\left\|\mathcal{L} - \left(\mathcal{M}^{k+\frac{1}{2}} + \frac{1}{\beta}\mathcal{Z}^k - \mathcal{E}^k\right)\right\|_F^2. \tag{20}$$

By Theorem 3, the minimizer of problem (20) is given by

$$\mathcal{L}^{k+1} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S}_\beta \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \tag{21}$$

where $\mathcal{M}^{k+\frac{1}{2}} + \frac{1}{\beta}\mathcal{Z}^k - \mathcal{E}^k = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H$ and $\mathcal{S}_\beta = \Phi^H[\hat{\mathcal{S}}_\beta]$ with $\hat{\mathcal{S}}_\beta = \max\{\hat{\mathcal{S}}_{\mathbf{\Phi}} - \frac{1}{\beta}, 0\}$. The minimizer with respect to $\mathcal{E}$ in (14) can be given by

$$\mathcal{E}^{k+1} = \text{sgn}(\mathcal{H}) \circ \max\left\{ |\mathcal{H}| - \frac{\lambda}{\beta}, 0 \right\}, \tag{22}$$

where $\mathcal{H} := \mathcal{M}^{k+1} + \frac{1}{\beta}\mathcal{Z}^k - \mathcal{L}^{k+1}$, $\circ$ denotes the pointwise product, and $\text{sgn}(\cdot)$ denotes the signum function, that is,

$$\text{sgn}(y) := \begin{cases} 1, & \text{if } y > 0, \\ 0, & \text{if } y = 0, \\ -1, & \text{if } y < 0. \end{cases}$$

Since only two blocks of the objective function in (10) are nonsmooth and the other block is not involved in Equation (10), the sGS-ADMM is convergent.[48,th.3] The sGS-ADMM for solving Equation (10) can be stated in Algorithm 2.

Algorithm 2. A symmetric Gauss-Seidel-based multiblock ADMM for solving (10)

Input: $\tau \in (0, (1+\sqrt{5})/2), \beta > 0, \mathcal{L}^0, \mathcal{E}^0, \mathcal{Z}^0$. For $k = 0, 1, 2, \ldots$, perform the following steps:
1: Compute $\mathcal{M}^{k+\frac{1}{2}}$ by

$$\mathcal{M}^{k+\frac{1}{2}} = \begin{cases} \mathcal{X}_{ijk}, & \text{if } (i,j,k) \in \Omega, \\ \left(\mathcal{L}^k + \mathcal{E}^k - \frac{1}{\beta}\mathcal{Z}^k\right)_{ijk}, & \text{otherwise.} \end{cases}$$

2: Compute $\mathcal{L}^{k+1}$ via (21).
3: Compute $\mathcal{M}^{k+1}$ by

$$\mathcal{M}^{k+1} = \begin{cases} \mathcal{X}_{ijk}, & \text{if } (i,j,k) \in \Omega, \\ \left(\mathcal{L}^{k+1} + \mathcal{E}^k - \frac{1}{\beta}\mathcal{Z}^k\right)_{ijk}, & \text{otherwise.} \end{cases}$$

4: Compute $\mathcal{E}^{k+1}$ via (22).
5: Update $\mathcal{Z}^{k+1}$ by (15).
6: If a stopping criterion is not met, set $k := k+1$ and go to step 1.

4 | EXPERIMENTAL RESULTS

In this section, numerical results are presented to demonstrate the effectiveness of the proposed method for robust tensor completion. We compare the transformed tensor SVD with the sum of nuclear norms of unfolding matrices of a tensor plus a sparse tensor (SNN)*,[52] tensor SVD using Fourier transform (t-SVD (Fourier))†,[45] and low-rank tensor completion by parallel matrix factorization (TMac)‡.[53] All the experiments are performed under Windows 7 and MATLAB R2018a running on a desktop (Intel Core i7, @ 3.40 GHz, 8.00 G RAM).

4.1 | Experimental setting

The sampling ratio of observations is defined as $\rho := \frac{\#\Omega}{n_1 n_2 n_3}$, where Ω is generated uniformly at random. In order to evaluate the performance of different methods for real-world tensors, the PSNR is used to measure the quality of the estimated tensors, which is defined as follows:

$$\text{PSNR} := 10\log_{10}\frac{n_1 n_2 n_3(\mathcal{L}_{\max} - \mathcal{L}_{\min})^2}{\|\mathcal{L} - \mathcal{L}_0\|_F^2},$$

where $\mathcal{L}$ is the recovered solution, $\mathcal{L}_0$ is the ground-truth tensor, $\mathcal{L}_{\max}$ and $\mathcal{L}_{\min}$ are maximal and minimal entries of $\mathcal{L}_0$, respectively.

As suggested by Theorem 2, we set $\lambda = \frac{a}{\sqrt{n_{(1)}n_3}}$ and adjust it slightly to obtain the best possible results. In all experiments, a is selected from $\{1.1, 1.3, 1.5, 1.7, 2\}$ in Fourier transform and from $\{10, 15, 18, 20, 23, 25, 28, 30, 33, 35, 40, 45, 50\}$ in unitary and wavelet transforms. The parameter λ in Fourier transform is different from that in unitary and wavelet transforms in each case, and for unitary and wavelet transforms, it is set to be the same in each case. Moreover, τ is set to be 1.618 for its convergence[48] and β is chosen from $\{0.01, 0.05, 0.1\}$ to get the highest PSNR values in Algorithm 2. The parameter β is the same across the unitary transform and Fourier transform methods.

*https://tonyzqin.wordpress.com/.
†https://canyilu.github.io/publications/.
‡https://xu-yangyang.github.io/TMac/.

The Karush-Kuhn-Tucker (KKT) conditions of problem (10) are given by

$$\begin{cases} \mathcal{Z} \in \partial\|\mathcal{L}\|_{\mathrm{TTNN}}, \quad \mathcal{Z} \in \partial\lambda\|\mathcal{E}\|_1, \\ \mathcal{L} + \mathcal{E} = \mathcal{M}, \quad \mathcal{P}_\Omega(\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \end{cases}$$

where $\partial\|\mathcal{L}\|_{\mathrm{TTNN}}$ and $\partial\lambda\|\mathcal{E}\|_1$ denote the subdifferentials of $\|\cdot\|_{\mathrm{TTNN}}$ and $\lambda\|\cdot\|_1$, respectively. Note that $\mathcal{P}_\Omega(\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X})$ is always satisfied in each iteration of the sGS-ADMM. We measure the accuracy of an approximate optimal solution by using the following relative KKT residual:

$$\eta_{\mathrm{res}} := \max\{\eta_z, \eta_e, \eta_m\},$$

where

$$\eta_z = \frac{\|\mathcal{L} - \mathrm{Prox}_{\|\cdot\|_{\mathrm{TTNN}}}(\mathcal{Z} + \mathcal{L})\|_F}{1 + \|\mathcal{Z}\|_F + \|\mathcal{L}\|_F}, \quad \eta_e = \frac{\|\mathcal{E} - \mathrm{Prox}_{\lambda\|\cdot\|_1}(\mathcal{Z} + \mathcal{E})\|_F}{1 + \|\mathcal{Z}\|_F + \|\mathcal{E}\|_F}, \quad \eta_m = \frac{\|\mathcal{L} + \mathcal{E} - \mathcal{M}\|_F}{1 + \|\mathcal{L}\|_F + \|\mathcal{E}\|_F + \|\mathcal{M}\|_F}.$$

Here Prox_g is the proximal mapping of g, that is, $\mathrm{Prox}_g(y) = \arg\min_x \left\{ g(x) + \frac{1}{2}\|x - y\|^2 \right\}$. The stopping criterion of the Algorithm is set to $\eta_{\mathrm{res}} \le 5 \times 10^{-4}$ and the maximum number of iterations is set to be 500.

For the sparse level of $\mathcal{E}$, a fraction γ of its entries are uniformly corrupted by additive independent and identically distributed noise from a standard Gaussian distribution $N(0, 1)$ at random, which generates the sparse tensor $\mathcal{E}$. The testing real-world tensors are rescaled in $[0, 1]$.

In theory, we discuss a general unitary transformed method in the tensor product and robust tensor completion problem in Section 3. In the following test, we consider two different kinds of unitary transformations in the proposed method.

- The first one is a Daubechies 4 (db4) discrete wavelet transform in the periodic mode[54] to compute the transformed tensor SVD (called t-SVD (wavelet)).
- The second one is based on given tensor data to construct a unitary transform matrix (called t-SVD (data)). For the incomplete tensor $\mathcal{P}_\Omega(\mathcal{X})$ in (5), we first give an initial estimator $\mathcal{A}$, which is generated by t-SVD (Fourier). Then we unfold $\mathcal{A}$ into a matrix $\mathbf{A}$ along the third-dimension, that is, $\mathbf{A} = \mathcal{A}_{(3)}$, where $\mathcal{A}_{(3)}$ denotes the mode-3 unfolding of $\mathcal{A}$. Let the SVD of the unfolding matrix $\mathbf{A}$ be $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. $\mathbf{U}^H$ is the desired unitary matrix in the $\mathbf{\Phi}$-product and transformed tensor SVD. Suppose that $\mathrm{rank}(\mathbf{A}) = r$. It is interesting to see that $\mathbf{U}^H$ is the optimal transformation to obtain a low rank matrix of $\mathbf{A}$:

$$\min_{\mathrm{rank}(\mathbf{B})=r,\ \mathrm{unitary}\ \mathbf{\Phi}} \|\mathbf{\Phi}\mathbf{A} - \mathbf{B}\|_F^2.$$

  Therefore, if the unfolding matrix $\mathbf{A}$ of the underlying tenor is low-rank, then $\mathbf{U}^H\mathbf{A}$ is also a low rank matrix and its rank is r. Denote $\mathrm{Fold}_3(\mathcal{A}_{(3)}) = \mathcal{A}$. Then we fold $\mathbf{U}^H\mathbf{A}$ into a tensor (denoted by $\mathrm{Fold}_3(\mathbf{U}^H\mathbf{A})$) and the last $(n_3 - r)$ frontal slices of the tensor $\mathrm{Fold}_3(\mathbf{U}^H\mathbf{A})$ are all zero matrices. Hence, the t-SVD (data) method may be useful since many frontal slices of the underlying tensor in the transformed domain are zero matrices, which implies its transformed multirank may be very low.

We consider the computational cost of t-SVD by different transformations. The application of discrete Fourier transform to an n_3-vector is of $O(n_3 \log n_3)$ operations. There are $n_1 n_2$ tubes in an $n_1 \times n_2 \times n_3$ tensor. Assume $n_2 \le n_1$. In the Fourier transformed tensor SVD, we need to compute n_3 n_1-by-n_2 SVDs and the cost is $O(n_1 n_2^2 n_3)$. The total cost is of $O(n_1 n_2^2 n_3 + n_1 n_2 n_3 \log n_3)$ operations. The application of Daubechies 4 (db4) discrete wavelet transform to an n_3-vector is of $O(n_3)$ operations. There are still n_3 n_1-by-n_2 SVDs to be calculated. The total cost is of $O(n_1 n_2^2 n_3 + n_1 n_2 n_3)$ operations. The application of a unitary transformation to an n_3-vector is of $O(n_3^2)$ operations. There are still n_3 n_1-by-n_2 SVDs to be calculated. The total cost is of $O(n_1 n_2^2 n_3 + n_1 n_2 n_3^2)$ operations. When n_2 is larger than n_3, then the computational cost of the above three transformations are about the same. However, when n_3 is larger than n_2, the computational cost for the use of unitary transformation is larger than that by using the Fourier transform and the wavelet transform. We note that in the following testing hyperspectral, video and face datasets, n_3 is larger than $\min(n_1, n_2)$. We find that the computational time required by the use of unitary transformation is larger than that required by the use of the Fourier transform.

However, we see that the error of the tensor recovery by the use of unitary transformation is better than that by the use of the Fourier transformation.

4.2 | Hyperspectral data

In this subsection, we use three hyperspectral datasets: Samson, Japser Ridge, and Urban datasets[55] to show the effectiveness of the proposed method. The testing datasets are third-order tensors (Length × Width × Channels). We describe the three datasets in the following:

- For the Samson dataset, we only utilize a region of 95×95 in each image, where each pixel is recorded at 156 frequency channels covering the wavelengths from 401 to 889 nm. Then the spectral resolution is highly up to 3.13 nm. Thus, the size of the resulting tensor is $95 \times 95 \times 156$.
- For the Japser Ridge dataset, each pixel is recorded at 224 frequency channels with wavelengths being from 380 to 2,500 nm. The spectral resolution is up to 9.46 nm. Since this hyperspectral image is too complex to get the ground truth, a subimage of 100×100 pixels is considered. The first pixel starts from the (105, 269)th pixel in the original image. Due to dense water vapor and atmospheric effects, we only remain 198 channels. Therefore, the size of the resulting tensor is $100 \times 100 \times 198$.
- For the Urban dataset, there are 307×307 pixels of each image, each of which corresponds to a $2 \times 2\,\mathrm{m}^2$ area. In this image, there are 210 wavelengths ranging from 400 to 2,500 nm, which results in a spectral resolution of 10 nm. One hundred and sixty two channels of this dataset are remained due to dense water vapor and atmospheric effects. Hence, the size of the resulting tensor is $307 \times 307 \times 162$.

First we show the distributions of all singular values of the testing hyperspectral datasets by different methods in the transformed domain. For the original tensor $\mathcal{L}_0 = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we show the distributions of the singular values $\hat{\mathcal{S}}_{\mathbf{\Phi}}(i,i,j)$ in different ranges, $1 \le i \le \mathrm{rank}(\mathcal{L}_0)$, $1 \le j \le n_3$. The distributions of all singular values $\hat{\mathcal{S}}_{\mathbf{\Phi}}(i,i,j)$ obtained by t-SVD (Fourier), t-SVD (wavelet), t-SVD (data) in different ranges are shown in Figure 1. It can be observed that the number of small singular values $\hat{\mathcal{S}}_{\mathbf{\Phi}}(i,i,j)$ obtained by t-SVD (data) is much larger than that obtained by t-SVD (Fourier) and t-SVD (wavelet). When discarding these small singular values, the resulting tensor may have low tubal rank since the number of large singular values is much smaller. In the robust tensor completion problem, we minimize the TTNN, and the recovered tensor can be achieved by truncating small tensor singular values and obtaining a lower tubal rank tensor.

We consider the robust tensor completion problem for the testing hyperspectral datasets with different sampling ratios and γ. Figure 2 displays the visual comparisons of different methods for the Japser Ridge dataset with 60% sampling ratio and 30% corruption entries. We can observe that the visual quality obtained by t-SVD (data) is better than that obtained by SNN, TMac, t-SVD (Fourier), and t-SVD (wavelet). The PSNR values obtained by different methods are displayed in Table 1. We can observe that the PSNR values obtained by t-SVD (data) are much higher than those obtained by SNN,

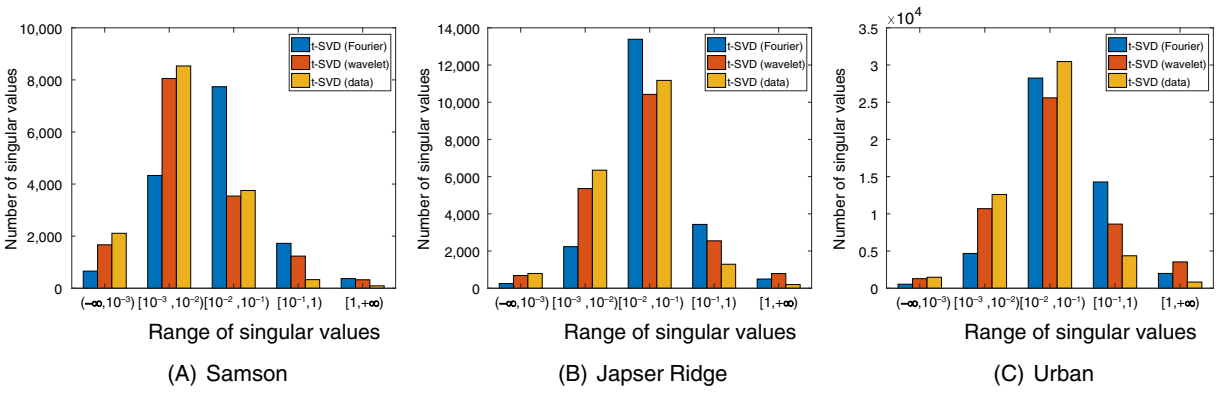

FIGURE 1 Distributions of singular values of all frontal slices in different ranges by t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the hyperspectral datasets. (A) Samson. (B) Japser Ridge. (C) Urban

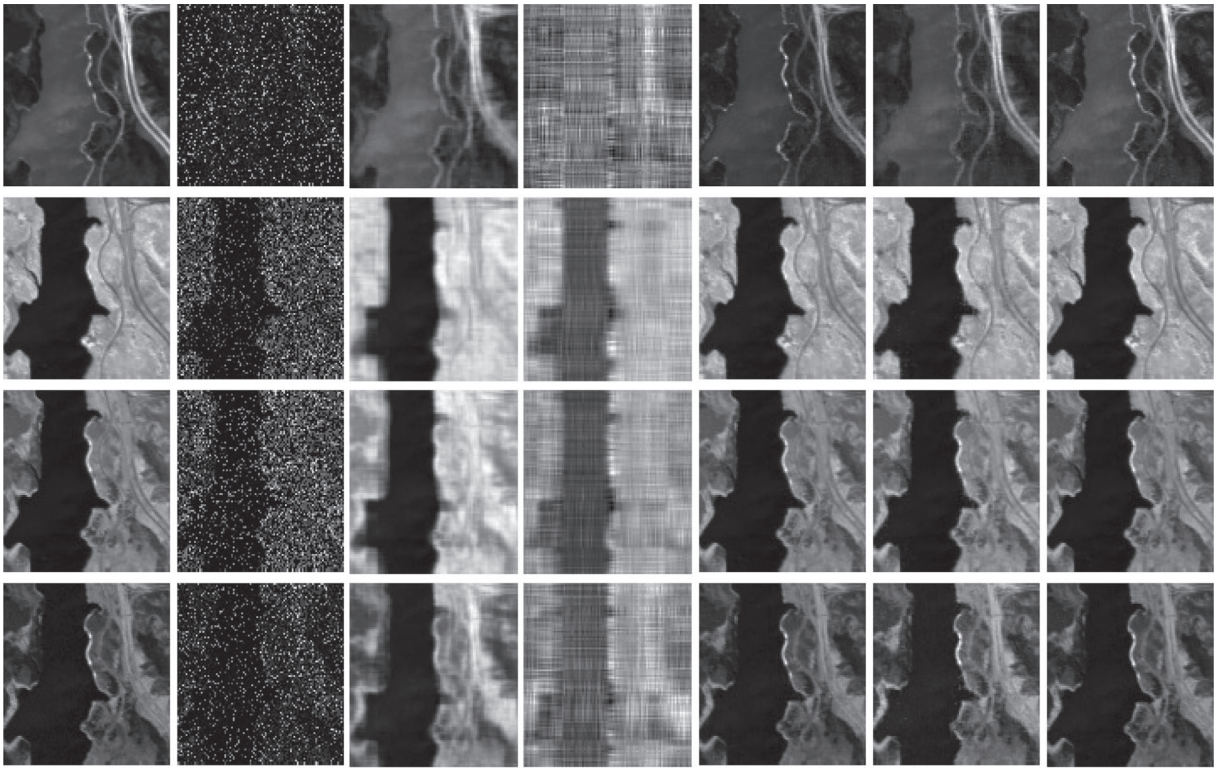

FIGURE 2 (A) Original images for the Japser Ridge dataset. (B) Observed images (60% sampling ratio and 30% corrupted entries). (C) Recovered images by SNN [PSNR = 26.00]. (D) Recovered images by TMac [PSNR = 16.47]. (E) Recovered images by t-SVD (Fourier) [PSNR = 33.63]. (F) Recovered images by t-SVD (wavelet) [PSNR = 33.59]. (G) Recovered images by t-SVD (data) [PSNR = 37.38]

TMac, t-SVD, and t-SVD (wavelet) for different sampling ratios $(0.6, 0.8)$ and γ $(0.1, 0.2, 0.3)$. The improvements of t-SVD (data) are very impressive, especially for small γ. The performance of t-SVD (wavelet) is better than that of SNN, TMac, and t-SVD (Fourier) in terms of PSNR values for the Samson and Japser Ridge datasets. For the Urban dataset, the PSNR values obtained by t-SVD (Fourier) are slightly higher than those obtained by t-SVD (wavelet), especially for large γ.

4.3 | Video data

In this subsection, we present three video data (Length × Width × Frames) including Carphone ($144 \times 176 \times 180$), Galleon ($144 \times 176 \times 200$), and Announcer ($144 \times 176 \times 200$)[§] to show the effectiveness of the proposed method in the robust tensor completion problem, where the first channels of all frames in the original data are used. We just choose 180 and 200 frames for these videos to improve the computational time.

Similar to the hyperspectral datasets, we generate all singular values of the video datatsets in the transformed domain by t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data). In Figure 3, the distributions of all singular values by t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) are shown in different ranges. It can be seen that the number of small singular values obtained by t-SVD (data) is larger than those obtained by t-SVD (Fourier) and t-SVD (data). Moreover, for the Carphone and Galleon datasets, the number of small singular values obtained by t-SVD (wavelet) is larger than those obtained by t-SVD (Fourier). Hence, the tubal ranks of the three video datasets obtained by t-SVD (data) are lower than those obtained by t-SVD (Fourier) and t-SVD (wavelet) when we discard some small singular values. Furthermore, the tubal ranks of the Carphone and Galleon datasets obtained by t-SVD (wavelet) are lower than those by t-SVD (Fourier) when discarding some small singular values.

[§]http://trace.eas.asu.edu/yuv/.

TABLE 1 PSNR values obtained by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the hyperspectral datasets

	ρ	γ	SNN	TMac	t-SVD (Fourier)	t-SVD (wavelet)	t-SVD (data)
Samson	0.6	0.1	32.39	23.15	38.53	45.43	**53.30**
		0.2	30.35	19.35	34.80	41.29	**50.68**
		0.3	28.50	16.82	32.26	38.22	**45.87**
	0.8	0.1	33.55	23.28	40.87	47.34	**54.40**
		0.2	31.69	19.42	36.82	44.69	**52.49**
		0.3	30.04	16.86	33.58	39.46	**48.28**
Japser Ridge	0.6	0.1	30.13	21.73	39.22	40.60	**45.08**
		0.2	27.92	18.76	36.38	37.20	**41.13**
		0.3	26.00	16.47	33.63	33.59	**37.38**
	0.8	0.1	31.98	21.84	40.78	42.64	**46.76**
		0.2	29.61	18.81	37.88	38.87	**43.13**
		0.3	27.49	16.51	35.15	35.81	**39.33**
Urban	0.6	0.1	27.88	22.20	38.78	39.10	**47.76**
		0.2	26.13	18.43	36.08	35.70	**44.51**
		0.3	24.69	16.06	33.39	32.16	**39.63**
	0.8	0.1	30.31	22.29	41.29	42.94	**50.07**
		0.2	27.89	18.45	37.77	38.44	**45.98**
		0.3	26.06	16.07	34.98	34.29	**42.26**

Note: The boldface number is the best and the underline number is the second best.

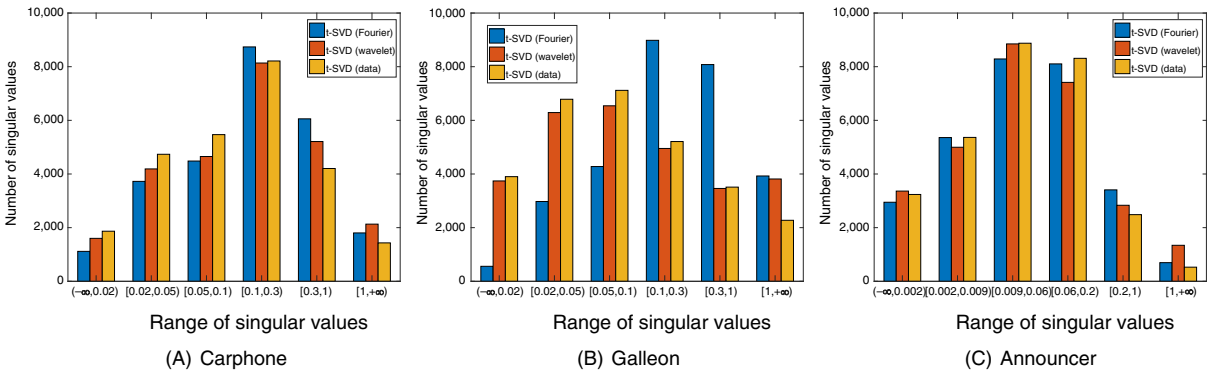

FIGURE 3 Distributions of singular values of all frontal slices in different ranges by t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the video datasets. (A) Carphone. (B) Galleon. (C) Announcer

We display the visual comparisons of the testing data in robust tensor completion with 60% sampling ratio and 20% corruption entries by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) in Figure 4. We can see that the images recovered by t-SVD (data) are better than those recovered by SNN, TMac, t-SVD (Fourier), and t-SVD (wavelet) in terms of visual quality. The t-SVD (data) can keep more details than SNN, TMac, t-SVD (Fourier), and t-SVD (wavelet) for the three testing videos.

We also show the PSNR values obtained by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the Carphone, Galleon, and Announcer datasets with different sampling ratios $(0.6, 0.8)$ and γ $(0.1, 0.1, 0.3)$ in Table 2. It can be seen that the PSNR values obtained by t-SVD (data) are higher than those obtained by SNN, TMac, t-SVD

FIGURE 4 Recovered images by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) in robust tensor completion for video data with 60% sampling ratio and 20% corrupted entries. (A) Original images. (B) Observed images. (C) SNN. (D) TMac. (E) t-SVD (Fourier). (F) t-SVD (wavelet). (G) t-SVD (data)

(Fourier), and t-SVD (wavelet). The PSNR values obtained by t-SVD (data) can be improved around 2 dB compared with those obtained by t-SVD (Fourier) for these datasets. For the Carphone and Galleon videos, the performance of t-SVD (wavelet) is better than that of t-SVD (Fourier) in terms of PSNR values. While the PSNR values obtain by t-SVD (Fourier) is slightly higher than those obtained by SNN, TMac, and t-SVD (wavelet) for large γ such as 0.2 and 0.3 cases.

4.4 | Face data

In this subsection, we use the extended Yale face database B[¶] to test the robust tensor completion problem. To improve the computational time, we crop the original image to contain the face and resize it to 73×55. Moreover, we only choose first 30 subjects and 25 illuminations in our test. Then the size of the testing tensor is $73 \times 55 \times 750$.

Being similar to the hyperspectral dataset, Figure 5 displays the distributions of singular values of all frontal slices of the extended Yale face database B in the transformed domain by t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data). We can see that the t-SVD (data) generates more small singular values than t-SVD (Fourier) and t-SVD (wavelet). Moreover, the number of small singular values obtained by t-SVD (wavelet) is larger than that obtained by t-SVD (Fourier). These imply that the t-SVD (data) may get better recovery performance than t-SVD (Fourier) and t-SVD (wavelet) since the tubal rank obtained by t-SVD (data) is lower than those obtained by the other two t-SVD methods when we discard some small singular values.

The visual comparisons of SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the extended Yale face database B are shown in Figure 6, where the sampling ratio is 0.6 and $\gamma = 0.2$. It can be seen that the images obtained by t-SVD (data) are more clear than those obtained by SNN, TMac, t-SVD (Fourier), and t-SVD (wavelet).

In Table 3, we show the PSNR values of different sampling ratios (0.6, 0.8) and γ (0.1, 0.2, 0.3) for the extended Yale face database B in the robust tensor completion. It can be seen that the PSNR values obtained by t-SVD (data) are higher than those obtained by SNN, TMac, t-SVD (Fourier), and t-SVD (wavelet) for these sampling ratios and γ. The PSNR values of t-SVD (data) can be improved around 2 dB than those of t-SVD (Fourier).

[¶]http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html.

TABLE 2 The PSNR values obtained by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the video datasets

	ρ	γ	SNN	TMac	t-SVD (Fourier)	t-SVD (wavelet)	t-SVD (data)
Carphone	0.6	0.1	26.80	20.86	30.70	31.21	**32.38**
		0.2	24.88	17.35	28.82	28.30	**30.14**
		0.3	23.39	15.19	27.21	27.32	**28.09**
	0.8	0.1	28.74	21.03	32.57	32.73	**34.06**
		0.2	26.35	17.47	30.13	30.25	**31.29**
		0.3	24.68	15.27	28.16	28.18	**29.10**
Galleon	0.6	0.1	24.56	20.47	27.44	29.18	**29.80**
		0.2	22.14	17.19	25.63	26.55	**27.55**
		0.3	19.07	15.12	24.05	24.26	**25.84**
	0.8	0.1	26.86	20.75	29.37	30.67	**31.71**
		0.2	24.32	17.37	27.01	28.44	**29.03**
		0.3	22.10	15.23	25.08	26.26	**26.93**
Announcer	0.6	0.1	29.58	21.14	37.90	38.24	**39.44**
		0.2	27.55	17.52	35.35	35.07	**36.57**
		0.3	26.36	15.33	32.64	31.94	**34.04**
	0.8	0.1	31.57	21.24	39.62	40.45	**41.28**
		0.2	28.67	17.58	36.61	35.91	**37.97**
		0.3	27.50	15.37	34.15	33.54	**35.23**

Note: The boldface number is the best and the underline number is the second best.

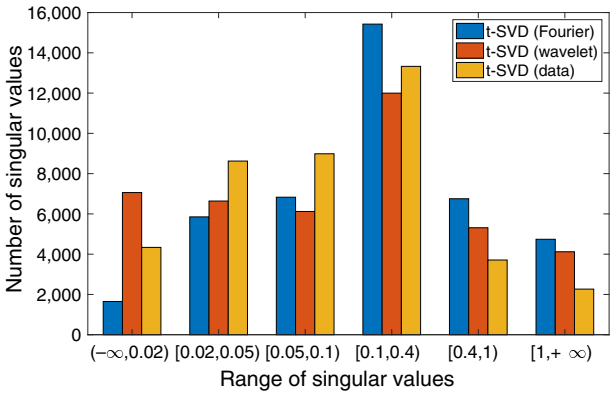

FIGURE 5 Distribution of singular values of all frontal slices in different ranges by t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the extended Yale face database B

5 | CONCLUDING REMARKS

We have studied the robust tensor completion problem by using transformed tensor SVD, which employs other unitary transform matrices instead of discrete Fourier transform matrix. The algebraic structure of the associated tensor product between two tensors is not necessary to be known in general and the tensor product can be defined via the unitary transformation directly. With this generalized tensor product, we have shown that one can recover a low transformed tubal rank tensor exactly with overwhelming probability provided that its transformed tubal rank is sufficiently small and its corrupted entries are reasonably sparse. Moreover, we have proposed an "optimal" data-dependent transform method in the robust tensor completion problem for third-order tensors. Numerical examples on many real-word tensors show the

FIGURE 6 Recovered images by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) in robust tensor completion for the extended Yale face database B with 60% sampling ratio and 20% corrupted entries. (A) Original images. (B) Observed images. (C) SNN. (D) TMac. (E) t-SVD (Fourier). (F) t-SVD (wavelet). (G) t-SVD (data)

TABLE 3 The PSNR values obtained by SNN, TMac, t-SVD (Fourier), t-SVD (wavelet), and t-SVD (data) for the extended Yale face database B

ρ	γ	SNN	TMac	t-SVD (Fourier)	t-SVD (wavelet)	t-SVD (data)
0.6	0.1	24.57	19.51	26.00	26.85	**28.66**
	0.2	22.89	17.60	24.06	24.73	**26.16**
	0.3	21.76	15.77	22.22	22.97	**23.75**
0.8	0.1	26.59	19.54	28.07	28.86	**30.37**
	0.2	24.53	17.79	25.54	26.02	**27.50**
	0.3	23.06	15.93	23.59	24.14	**25.19**

Note: The boldface number is the best and the underline number is the second best.

usefulness of the transformed tensor SVD method with wavelet and data-dependent transformations, and demonstrate that the performance of the proposed method is better than that of existing tensor completion methods.

ACKNOWLEDGEMENTS

The authors would like to thank the two anonymous referees for their valuable comments and remarks that have helped to improve the quality of the article greatly.

Research supported in part by the HKRGC GRF 12306616, 12200317, 12300218 and 12300519, and HKU Grant 104005583.

Research supported in part by the National Natural Science Foundation of China under grants 11801206, 11571098, 11871025, Hubei Provincial Natural Science Foundation of China under grant 2018CFB105, and Fundamental Research Funds for the Central Universities under grant CCNU19ZN017.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Michael K. Ng https://orcid.org/0000-0001-6833-5227
Xiongjun Zhang https://orcid.org/0000-0002-5387-4129

REFERENCES

1. Liu J, Musialski P, Wonka P, Ye J. Tensor completion for estimating missing values in visual data. IEEE Trans Pattern Anal Mach Intell. 2013;35(1):208–220.
2. Nguyen TD, Lee G. Color image segmentation using tensor voting based color clustering. Pattern Recogn Lett. 2012;33(5):605–614.
3. Plataniotis KN, Venetsanopoulos AN. Color Image Processing and Applications. Berlin, Heidelberg / Germany: Springer, 2000.
4. Fan H, Li J, Yuan Q, Liu X, Mg MK. Hyperspectral image denoising with bilinear low rank matrix factorization. Signal Process. 2019;163:132–152.
5. Ng MK, Yuan Q, Yan L, Sun J. An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data. IEEE Trans Geosci Remote Sens. 2017;55(6):3367–3381.
6. Yang J-H, Zhao X-L, Ma T-H, Chen Y, Huang T-Z, Ding M. Remote sensing image destriping using unidirectional high-order total variation and nonconvex low-rank regularization. J Comput Appl Math. 2020;363:124–144.
7. Cichocki A, Mandic D, De Lathauwer L, et al. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Process Mag. 2015;32(2):145–163.
8. Cui L-B, Li M-H, Song Y. Preconditioned tensor splitting iterations method for solving multi-linear systems. Appl Math Lett. 2019;96:89–94.
9. Ding M, Huang T-Z, Ji T-Y, Zhao X-L, Yang J-H. Low-rank tensor completion using matrix factorization based on tensor train rank and total variation. J Sci Comput. 2019;81(2):941–964.
10. KoldaT. G. and Sun J. Scalable tensor decompositions for multi-aspect data mining. Paper presented at: Proceedings of the 8th International Conference on Data Mining; 2008. p. 363–372.
11. Miwakeichi F, Valdes-Sosa PA, Aubert-Vazquez E, et al. Decomposing EEG data into space-time-frequency components using parallel factor analysis and its relation with cerebral blood flow. Paper presented at: Proceedings of the International Conference on Neural Information Processing; 2007. pages 802–810.
12. Omberg L, Golub GH, Alter O. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. Proc Natl Acad Sci. 2007;104(47):18371–18376.
13. Rabanser S, Shchur O, Günnemann S. Introduction to tensor decompositions and their applications in machine learning. arXiv:1711.10781; 2017.
14. Xie Q, Zhao Q, Meng D, Xu Z. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. IEEE Trans Pattern Anal Mach Intell. 2018;40(8):1888–1902.
15. Zhang X. A nonconvex relaxation approach to low-rank tensor completion. IEEE Trans Neural Netw Learn Syst. 2019;30(6):1659–1671.
16. Fazel M. Matrix rank minimization with applications [PhD thesis]. Stanford University; 2002.
17. Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 2010;52(3):471–501.
18. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? J ACM. 2011;58(3):11.
19. Candès EJ, Recht B. Exact matrix completion via convex optimization. Found Comput Math. 2009;9(6):717–772.
20. Chen Y. Incoherence-optimal matrix completion. IEEE Trans Inform Theory. 2015;61(5):2909–2923.
21. Oymak S, Jalali A, Fazel M, Eldar YC, Hassibi B. Simultaneously structured models with application to sparse and low-rank matrices. IEEE Trans Inform Theory. 2015;61(5):2886–2908.
22. Recht B. A simpler approach to matrix completion. J Mach Learn Res. 2009;12:3413–3430.
23. Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. Stud Appl Math. 1927;6(1-4):164–189.
24. Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Rev. 2009;51(3):455–500.
25. Jain P, Oh S. Provable tensor factorization Q4 with missing data. Paper presented at: Proceedings of the Neural Information Processing Systems; 2014. p. 1431-1439.
26. Karlsson L, Kressner D, Uschmajew A. Parallel algorithms for tensor completion in the CP format. Parallel Comput. 2016;57:222–234.
27. Oseledets IV. Tensor-train decomposition. SIAM J Sci Comput. 2011;33(5):2295–2317.
28. Bengua JA, Phien HN, Tuan HD, Do MN. Efficient tensor completion for color image and video recovery: Low-rank tensor train. IEEE Trans Image Process. 2017;26(5):2466–2479.
29. Tucker LR. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966;31(3):279–311.
30. Romera-Paredes B, Pontil M. A new convex relaxation for tensor completion. Paper presented at: Proceedings of the Neural Information Processing Systems; 2013. p. 2967-2975.
31. Mu C, Huang B, Wright J, and Goldfarb D. Square deal: Lower bounds and improved relaxations for tensor recovery. In *Proc. Int. Conf. Mach. Learn.*, volume 32, pages 73–81, 2014.
32. Gandy S, Recht B, Yamada I. Tensor completion and low-n-rank tensor recovery via convex optimization. Inverse Probl. 2011;27(2):025010.
33. Gu Q, Gui H, Han J. Robust tensor decomposition with gross corruption. Paper presented at: Proceedings of the Neural Information Processing Systems; 2014. p. 1422-1430.
34. Huang B, Mu C, Goldfarb D, Wright J. Provable models for robust low-rank tensor completion. Pac J Optim. 2015;11(2):339–364.
35. Kilmer ME, Martin CD. Factorization strategies for third-order tensors. Linear Algebra Appl. 2011;435(3):641–658.
36. Kernfeld E, Kilmer M, Aeron S. Tensor–tensor products with invertible linear transforms. Linear Algebra Appl. 2015;485:545–570.
37. Martin CD, Shafer R, LaRue B. An order-p tensor factorization with applications in imaging. SIAM J Sci Comput. 2013;35(1):A474–A490.
38. Kilmer ME, Braman K, Hao N, Hoover RC. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM J Matrix Anal Appl. 2013;34(1):148–172.

39. Zhang Z, Ely G, Aeron S, Hao N, and Kilmer M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. Paper presented at: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2014. p. 3842–3849.
40. Zhang X, Ng MK. A corrected tensor nuclear norm minimization method for noisy low-rank tensor completion. SIAM J Imaging Sci. 2019;12(2):1231–1273.
41. Zhang Z, Aeron S. Exact tensor completion using t-SVD. IEEE Trans Signal Process. 2017;65(6):1511–1526.
42. Zhou P, Lu C, Lin Z, Zhang C. Tensor factorization for low-rank tensor completion. IEEE Trans Image Process. 2018;27(3):1152–1163.
43. Hu W, Tao D, Zhang W, Xie Y, Yang Y. The twist tensor nuclear norm for video completion. IEEE Trans Neural Netw Learn Syst. 2017;28(12):2961–2973.
44. Jiang Q, Ng MK. Robust low-tubal-rank tensor completion via convex optimization. Paper presented at: Proceedings of the 28th International Joint Conferences on Artificial Intelligence; 2019. p. 2649–2655.
45. Lu C, Feng J, Chen Y, Liu W, Lin Z, Yan S. Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans Pattern Anal Mach Intell. 2020;42(4):925-938.
46. Gross D. Recovering low-rank matrices from few coefficients in any basis. IEEE Trans Inform Theory. 2011;57(3):1548–1566.
47. Chen L, Sun D, Toh K-C. An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming. Math Progr. 2017;161(1-2):237–270.
48. Li X, Sun D, Toh K-C. A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions. Math Progr. 2016;155(1-2):333–373.
49. Bai M, Zhang X, Ni G, Cui C. An adaptive correction approach for tensor completion. SIAM J Imag Sci. 2016;9(3):1298–1323.
50. Wang B, Zou H. Another look at distance-weighted discrimination. J Royal Stat Soc B. 2018;80(1):177–198.
51. Cai J-F, Candès EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM J Optim. 2010;20(4):1956–1982.
52. Goldfarb D, Qin Z. Robust low-rank tensor recovery: Models and algorithms. SIAM J Matrix Anal Appl. 2014;35(1):225–253.
53. Xu Y, Hao R, Yin W, Su Z. Parallel matrix factorization for low-rank tensor completion. Inverse Probl Imag. 2015;9(2):601–624.
54. Daubechies I. Ten Lectures on Wavelets. Philadelphia, PA: SIAM, 1992.
55. Zhu F, Wang Y, Fan B, Xiang S, Meng G, Pan C. Spectral unmixing via data-guided sparsity. IEEE Trans Imag Process. 2014;23(12):5412–5427.

How to cite this article: Song G, Ng MK, Zhang X. Robust tensor completion using transformed tensor singular value decomposition. *Numer Linear Algebra Appl*. 2020;e2299. https://doi.org/10.1002/nla.2299

APPENDIX A. PROOF OF LEMMA 1

For convenience, we denote $\Upsilon(\mathcal{X}) = \text{rank}_{\text{sum}}(\mathcal{X})$. If the spectral norm of $\mathcal{X}$ is less than or equal to 1, the conjugate of the transformed multirank function $\Upsilon(\mathcal{X})$ on a unit ball of the tensor spectral norm can be defined as

$$\Upsilon^{\sharp}(\mathcal{Y}) = \sup_{\|\mathcal{X}\|_{\mathbf{\Phi}} \le 1} \left(\text{Re}(\langle \mathcal{Y}, \mathcal{X} \rangle) - \text{rank}_{\text{sum}}(\mathcal{X})\right),$$

where $\text{Re}(\cdot)$ denotes the real part of a complex number. Then by von Neumann's theorem and the tensor inner product given in Definition 5, we can get

$$\text{Re}(\langle \mathcal{Y}, \mathcal{X} \rangle) = \text{Re}(\langle \hat{y}_{\mathbf{\Phi}}, \hat{\mathcal{X}}_{\mathbf{\Phi}} \rangle) = \sum_{i=1}^{n_3} \text{Re}\left(\text{Tr}\left((\hat{y}_{\mathbf{\Phi}}^{(i)})^H \times \hat{\mathcal{X}}_{\mathbf{\Phi}}^{(i)}\right)\right) \le \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{X}_{\mathbf{\Phi}}}\right), \tag{A1}$$

where $\sigma_i\left(\overline{\mathcal{X}_{\mathbf{\Phi}}}\right)$ denotes the ith largest singular value of $\overline{\mathcal{X}_{\mathbf{\Phi}}}$. Let $\mathcal{Y} = \mathcal{U}_{\mathcal{Y}} \diamond_{\mathbf{\Phi}} \Sigma_{\mathcal{Y}} \diamond_{\mathbf{\Phi}} \mathcal{V}_{\mathcal{Y}}^H$ and $\mathcal{X} = \mathcal{U}_{\mathcal{X}} \diamond_{\mathbf{\Phi}} \Sigma_{\mathcal{X}} \diamond_{\mathbf{\Phi}} \mathcal{V}_{\mathcal{X}}^H$. Since $\|\mathcal{X}\|_{\mathbf{\Phi}} \le 1$, we can choose $\mathcal{U}_{\mathcal{X}} = \mathcal{U}_{\mathcal{Y}}$ and $\mathcal{V}_{\mathcal{X}} = \mathcal{V}_{\mathcal{Y}}$. Thus

$$\begin{aligned}
\text{Re}(\langle \mathcal{Y}, \mathcal{X} \rangle) &= \sum_{i=1}^{n_3} \text{Re}\left(\text{Tr}((\hat{y}_{\mathbf{\Phi}}^{(i)})^H \times \hat{\mathcal{X}}_{\mathbf{\Phi}}^{(i)})\right) \\
&= \text{Re}\left(\text{Tr}\left(\overline{(\mathcal{V}_{\mathcal{Y}})_{\mathbf{\Phi}}} \times \overline{(\Sigma_{\mathcal{Y}})_{\mathbf{\Phi}}} \times \overline{(\mathcal{U}_{\mathcal{Y}})_{\mathbf{\Phi}}^H} \times \overline{(\mathcal{U}_{\mathcal{Y}})_{\mathbf{\Phi}}} \times \overline{(\Sigma_{\mathcal{X}})_{\mathbf{\Phi}}} \times \overline{(\mathcal{V}_{\mathcal{Y}})_{\mathbf{\Phi}}^H}\right)\right) \\
&= \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{X}_{\mathbf{\Phi}}}\right),
\end{aligned}$$

which shows that the equality in (A1) can be obtained. Therefore, the conjugate of the transformed multirank function can be rewritten as

$$\Upsilon^{\sharp}(Y) = \sup_{\|\mathcal{X}\|_{\mathbf{\Phi}} \le 1} \left(\sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{X}_{\mathbf{\Phi}}}\right) - \mathrm{rank}_{\mathrm{sum}}(\mathcal{X}) \right). \tag{A2}$$

Now we first consider separated cases with the $\mathrm{rank}_{\mathrm{sum}}(\mathcal{X})$ being from 0 to $n_{(2)}n_3$. If $\mathrm{rank}_{\mathrm{sum}}(\mathcal{X}) = 0$, then $\Upsilon^{\sharp}(\mathcal{Y}) = 0$ for all $\mathcal{Y}$. If $\mathrm{rank}_{\mathrm{sum}}(\mathcal{X}) = r$, then

$$\Upsilon^{\sharp}(\mathcal{Y}) = \max\left\{ 0, \sigma_1\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1, \dots, \sum_{i=1}^{r} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - r, \dots, \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - n_{(2)}n_3 \right\},$$

for $1 \le r \le n_{(2)}n_3$. Furthermore, if $\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1 > 0$, the right hand of (A2) would be larger. Thus

$$\Upsilon^{\sharp}(\mathcal{Y}) = \begin{cases} 0, & \text{if } \left\|\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right\| \le 1, \\ \sum_{i=1}^{q} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - q, & \text{if } \sigma_q\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) > 1 \text{ and } \sigma_{q+1}\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) < 1,\ 1 \le q \le r. \end{cases}$$

The conjugate of $\Upsilon^{\sharp}(\mathcal{Y})$ is defined as

$$\Upsilon^{\sharp\sharp}(\mathcal{Z}) = \sup_{\mathcal{Y}} \left(\mathrm{Re}(\langle \mathcal{Z}, \mathcal{Y} \rangle) - \Upsilon^{\sharp}(\mathcal{Y}) \right) = \sup_{\overline{\mathcal{Y}_{\mathbf{\Phi}}}} \left(\mathrm{Tr}\left(\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right)^{H} \overline{\mathcal{Y}_{\mathbf{\Phi}}} \right) - \Upsilon^{\sharp}\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) \right),$$

for all $\mathcal{Z} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$. In a similar vein, suppose that $\mathcal{Z} = \mathcal{U}_{\mathcal{Z}} \diamond_{\mathbf{\Phi}} \Sigma_{\mathcal{Z}} \diamond_{\mathbf{\Phi}} \mathcal{V}_{\mathcal{Z}}^{H}$ and $\mathcal{Y} = \mathcal{U}_{\mathcal{Y}} \diamond_{\mathbf{\Phi}} \Sigma_{\mathcal{Y}} \diamond_{\mathbf{\Phi}} \mathcal{V}_{\mathcal{Y}}^{H}$, by choosing $\mathcal{U}_{\mathcal{Y}} = \mathcal{U}_{\mathcal{Z}}$ and $\mathcal{V}_{\mathcal{Y}} = \mathcal{V}_{\mathcal{Z}}$, we obtain that

$$\Upsilon^{\sharp\sharp}(\mathcal{Z}) = \sup_{\mathcal{Y}} \left(\sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - \Upsilon^{\sharp}(\mathcal{Y}) \right).$$

In the following, we can consider two cases, that is, $\left\|\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right\| \ge 1$ and $\left\|\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right\| \le 1$. In the first case, if $\left\|\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right\| \ge 1$, then $\sigma_1\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right)$ can be chosen large enough such that

$$\begin{aligned} \Upsilon^{\sharp\sharp}(\mathcal{Z}) &= \sup_{\mathcal{Y}} \left(\sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - \Upsilon^{\sharp}(\mathcal{Y}) \right) \\ &= \sup_{\mathcal{Y}} \left(\sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - \left(\sum_{i=1}^{q} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - q \right) \right) \\ &= \sup_{\mathcal{Y}} \left(\sigma_1\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) \left(\sigma_1\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) - 1 \right) + \left(\sum_{i=2}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - \left(\sum_{i=2}^{q} \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - q \right) \right) \right), \end{aligned}$$

tends to infinity. In the second case, $\left\|\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right\| \le 1$, if $\left\|\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right\| \le 1$, we obtain that $\Upsilon^{\sharp}(\mathcal{Y}) = 0$ and the supremum is achieved at $\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) = 1, i = 1, \dots, n_{(2)}n_3$, which yields

$$\Upsilon^{\sharp\sharp}(\mathcal{Z}) = \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) = \|\mathcal{Z}\|_{\mathrm{TTNN}}.$$

If $\left\|\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right\| > 1$, we can prove $\Upsilon^{\sharp\sharp}(\mathcal{Z}) \le \|\mathcal{Z}\|_{\mathrm{TTNN}}$. In fact,

$$\begin{aligned} &\sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - \sum_{i=1}^{q} \left(\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1 \right) \\ &= \sum_{i=1}^{q} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) + \sum_{i=q+1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - \sum_{i=1}^{q} \left(\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1 \right) - \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) + \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) \end{aligned}$$

$$= \sum_{i=1}^{q} \left(\sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) - 1\right)\left(\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1\right) + \sum_{i=q+1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right)\left(\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1\right) + \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right)$$
$$\leq \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right),$$

which can be derived from

$$\sum_{i=1}^{q} \left(\sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) - 1\right)\left(\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1\right) \leq 0 \quad \text{and} \quad \sum_{i=q+1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right)\left(\sigma_i\left(\overline{\mathcal{Y}_{\mathbf{\Phi}}}\right) - 1\right) \leq 0.$$

In summary, we can get

$$\Upsilon^{\#\#}(\mathcal{Z}) = \sum_{i=1}^{n_{(2)}n_3} \sigma_i\left(\overline{\mathcal{Z}_{\mathbf{\Phi}}}\right) = \|\mathcal{Z}\|_{\mathrm{TTNN}},$$

over the set $\|\mathcal{Z}\|_{\mathbf{\Phi}} \leq 1$.

In addition, the convex envelope of $\mathrm{rank}_{\mathrm{sum}}(\mathcal{X})$ on $\{\mathcal{X} \in \mathbb{C}^{n_1\times n_2\times n_3} \mid \|\mathcal{X}\|_{\mathbf{\Phi}} \leq c\}$ can be changed into $\mathrm{rank}_{\mathrm{sum}}(\mathcal{Y})$ on $\{\mathcal{Y} \in \mathbb{C}^{n_1\times n_2\times n_3} \mid \|\mathcal{Y}\|_{\mathbf{\Phi}} \leq 1\}$ by setting $\mathcal{Y} = \frac{1}{c}\mathcal{X}$.

APPENDIX B. PROOF OF THEOREM 2

In this appendix, we provide the detailed proof of Theorem 2. The idea is to employ convex analysis to derive the conditions in which one can check whether the pair $(\mathcal{L}_0, \mathcal{E}_0)$ is the unique minimizer to (5), and to show that such conditions are met with overwhelming probability in the conditions of Theorem 2. Before giving the detailed proof, we need to introduce the sampling schemes, some useful lemmas, as well as the subgradients of the tensor ℓ_1-norm and the TTNN used in Theorem 2, respectively.

Sampling schemes

The sampling strategy used in Theorem 2 is the uniform sampling without replacement. There are other widely used sampling models, for example, Bernoulli sampling, adaptive sampling, and random sampling with replacement. To facilitate our proof, we will consider the independent and identically distributed *(i.i.d.) Bernoulli-Rademacher model*. More precisely, we assume $\Omega = \{(i,j,k)|\delta_{ijk} = 1\}$, where δ_{ijk} are i.i.d. Bernoulli variables taking value one with probability ρ and zero with probability $1-\rho$. Such a Bernoulli sampling is denoted by $\Omega \sim \mathrm{Ber}(\rho)$ for short. As a proxy for uniform sampling, the probability of failure under Bernoulli sampling with $\rho = \frac{m}{n_1 n_2 n_3}$ closely approximates the probability of failure under uniform sampling.

Let a subset $\Lambda \subset \Omega$ be the corrupted entries of $\mathcal{L}_0$ and $\Gamma \subset \Omega$ be the locations where the data are available and clean. In a standard Bernoulli model, we suppose that

$$\Omega \sim \mathrm{Ber}(\rho), \quad \Lambda \sim \mathrm{Ber}(\gamma\rho), \quad \Gamma \sim \mathrm{Ber}((1-\gamma)\rho),$$

and the signs of the nonzero entries of $\mathcal{E}_0$ are deterministic. It has been shown to be much easier to work with a stronger assumption that the signs of the nonzero entries of $\mathcal{E}_0$ are independent symmetric ± 1 random variables (i.e., Rademacher random variables). We introduce two independent random subsets of Ω

$$\Lambda' \sim \mathrm{Ber}(2\gamma\rho), \quad \Gamma' \sim \mathrm{Ber}((1-2\gamma)\rho),$$

and it is convenient to consider $\mathcal{E}_0 = \mathcal{P}_\Gamma(\mathcal{E})$ for some fixed tensor $\mathcal{E}$. Assume that a random sign tensor $\mathcal{M}$ has i.i.d. entries such that for any index (i,j,k), $\mathbb{P}(\mathcal{M}_{i,j,k} = 1) = \mathbb{P}(\mathcal{M}_{i,j,k} = -1) = \frac{1}{2}$. Then $|\mathcal{E}|\circ\mathcal{M}$ has components with symmetric random signs. By introducing a new noise tensor $\mathcal{E}_0' = \mathcal{P}_{\Gamma'}(|\mathcal{E}|\circ\mathcal{M})$ and using the standard derandomization theory,[18,th.2.3] we have the following results.

Lemma 2. *Suppose that $\mathcal{L}_0$ obeys the conditions in Theorem 2 and $\mathcal{E}_0, \mathcal{E}_0'$ are given as above. If the recovery of $(\mathcal{L}_0, \mathcal{E}_0')$ is exact with high probability, it is also exact with at least the same probability for the model with the input data $(\mathcal{L}_0, \mathcal{E}_0)$.*

According to Lemma 2, we can equivalently assume that the nonzero entries have symmetric random signs and

$$\Lambda \sim \mathrm{Ber}(2\gamma\rho), \quad \Gamma \sim \mathrm{Ber}((1-2\gamma)\rho), \tag{B1}$$

for the locations of nonzero and zero entries of $\mathcal{E}_0$, respectively.

Useful lemmas

Suppose that $\mathcal{L}_0$ satisfies

$$\mathcal{L}_0 = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{S} \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \tag{B2}$$

where $\mathcal{U} \in \mathbb{C}^{n_1 \times r \times n_3}$, $\mathcal{V} \in \mathbb{C}^{n_2 \times r \times n_3}$ satisfying $\mathcal{U}^H \diamond_{\mathbf{\Phi}} \mathcal{U} = \mathcal{I}_{\mathbf{\Phi}}$, $\mathcal{V}^H \diamond_{\mathbf{\Phi}} \mathcal{V} = \mathcal{I}_{\mathbf{\Phi}}$, and $\mathcal{S} \in \mathbb{C}^{r \times r \times n_3}$ is a diagonal tensor. Denote the set T by

$$T = \left\{ \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{Y}^H + \mathcal{W} \diamond_{\mathbf{\Phi}} \mathcal{V}^H \mid \mathcal{Y} \in \mathbb{C}^{n_2 \times r \times n_3},\ \mathcal{W} \in \mathbb{C}^{n_1 \times r \times n_3} \right\}. \tag{B3}$$

For any $\mathcal{Z} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, the projections onto T and its complementary set are given as follows [40,prop.7.1]:

$$\begin{aligned} \mathcal{P}_T(\mathcal{Z}) &= \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{U}^H \diamond_{\mathbf{\Phi}} \mathcal{Z} + \mathcal{Z} \diamond_{\mathbf{\Phi}} \mathcal{V} \diamond_{\mathbf{\Phi}} \mathcal{V}^H - \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{U}^H \diamond_{\mathbf{\Phi}} \mathcal{Z} \diamond_{\mathbf{\Phi}} \mathcal{V} \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \\ \mathcal{P}_{T^\perp}(\mathcal{Z}) &= (\mathcal{I}_{\mathbf{\Phi}} - \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{U}^H) \diamond_{\mathbf{\Phi}} \mathcal{Z} \diamond_{\mathbf{\Phi}} (\mathcal{I}_{\mathbf{\Phi}} - \mathcal{V} \diamond_{\mathbf{\Phi}} \mathcal{V}^H). \end{aligned}$$

From the definition of $\mathcal{P}_T(\mathcal{A})$, and recall the inner product between two tensors, we can derive the following results.

Lemma 3. *Suppose that T is defined as (B3), then* $\langle \mathcal{P}_T(\mathcal{A}), \mathcal{B} \rangle = \langle \mathcal{A}, \mathcal{P}_T(\mathcal{B}) \rangle$.

It follows from Lemma 3 that $\langle \mathcal{P}_T(\mathcal{A}), \mathcal{P}_{T^\perp}(\mathcal{B}) \rangle = 0$, where $\mathcal{A}$ and $\mathcal{B}$ are arbitrary tensors with proper sizes. Furthermore, based on the transformed tensor incoherence conditions given in Equations (6) and (7), we can prove the following results which will be used many times in the sequel.

Lemma 4. *Let* $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ *be an arbitrary tensor, and T be given as (B3). Suppose that the transformed tensor incoherence conditions (6) and (7) are satisfied and denote* $\mathcal{E}_{ijk} = \vec{\boldsymbol{e}}_{ik} \diamond_{\mathbf{\Phi}} \mathring{\boldsymbol{e}}_{kk} \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{jk}^H$, *then*

$$\|\mathcal{P}_T(\vec{\boldsymbol{e}}_{ik} \diamond_{\mathbf{\Phi}} \mathring{\boldsymbol{e}}_{kk} \diamond_{\mathbf{\Phi}} \vec{\boldsymbol{e}}_{jk}^H)\|_F^2 = \|\mathcal{P}_T(\mathcal{E}_{ijk})\|_F^2 \le \frac{2\mu_0 r}{n_{(2)}}.$$

For any given tensor $\mathcal{Z} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, we have

$$\mathcal{P}_T(\mathcal{Z}) = \sum_{i,j,k} \left\langle \mathcal{P}_T(\mathcal{Z}), \mathcal{E}_{ijk} \right\rangle \mathcal{E}_{ijk} = \sum_{i,j,k} \left\langle \mathcal{Z}, \mathcal{P}_T(\mathcal{E}_{ijk}) \right\rangle \mathcal{E}_{ijk}. \tag{B4}$$

Moreover, setting $\Omega \sim \mathrm{Ber}(\rho)$, then

$$\rho^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T(\mathcal{Z}) = \rho^{-1} \sum_{i,j,k} \delta_{ijk} \left\langle \mathcal{Z}, \mathcal{P}_T(\mathcal{E}_{ijk}) \right\rangle \mathcal{P}_T(\mathcal{E}_{ijk}). \tag{B5}$$

Based on the tensor decomposition forms given in (B4) and (B5), we can get the following three lemmas by applying the methods used in Reference 44.

Lemma 5. *Suppose that* $\Omega \sim Ber(\rho)$, *and T is defined in (B3). Then with high probability,*

$$\|\rho^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T - \mathcal{P}_T\|_{op} \le \epsilon,$$

provided that $\rho \ge C_0 \frac{\mu r \log(n_{(1)} n_3)}{n_{(2)}} \epsilon^{-2}$ *for some numerical constant* $C_0 > 0$.

Lemma 6. *Suppose that* $\mathcal{Z} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ *is a fixed tensor and* $\Omega \sim \mathrm{Ber}(\rho)$. *Then with high probability,*

$$\|(\rho^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T - \mathcal{P}_T)\mathcal{Z}\|_\infty \le \epsilon \|\mathcal{Z}\|_\infty, \tag{B6}$$

provided that $\rho \geq C_0 \frac{\mu r \log(n_{(1)} n_3)}{n_{(2)}} \epsilon^{-2}$ *for some numerical constant* $C_0 > 0$.

Lemma 7. *Suppose that* $\mathcal{Z} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ *is a fixed tensor and* $\Omega \sim \text{Ber}(\rho)$. *Then with high probability,*

$$\|(\mathcal{I}_{\mathbf{\Phi}} - \rho^{-1}\mathcal{P}_{\Omega})\mathcal{Z}\|_{\mathbf{\Phi}} \leq C_0' \sqrt{\frac{n_{(1)} n_3 \log(n_{(1)} n_3)}{\rho}} \|\mathcal{Z}\|_{\infty}, \tag{B7}$$

provided that $\rho \geq C_0 \frac{\log(n_{(1)} n_3)}{n_{(2)} n_3}$ *for some numerical constants* $C_0, C0\prime > 0$.

The following lemma is similar to lemma 4.1 in Reference 45, which provides an upper bound for the spectral norm (which is based on any unitary transform instead of discrete Fourier transform) of the tensors consisting of Bernoulli sign variables. For simplicity, we omit its proof.

Lemma 8. *For the* $n_1 \times n_2 \times n_3$ *Bernoulli sign tensor* $\mathcal{M}$ *whose entries are distributed as*

$$\mathcal{M}_{ijk} = \begin{cases} 1, & \text{with probability } \frac{\rho}{2}, \\ 0, & \text{with probability } 1 - \rho, \\ -1, & \text{with probability } \frac{\rho}{2}, \end{cases} \tag{B8}$$

there exists a function $\varphi(\rho)$ *satisfying* $\lim_{\rho \to 0^+} \varphi(\rho) = 0$, *such that the following statement holds with high probability:*

$$\|\mathcal{M}\|_{\mathbf{\Phi}} \leq \varphi(\rho) \sqrt{n_{(1)} n_3}. \tag{B9}$$

Subgradients of the tensor ℓ_1-norm and the TTNN

In the optimization model (5), the tensor ℓ_1-norm and the TTNN are used. One of the main technical tools in analyzing the tensor norms minimization is the characterization of the subgradients. Then we list the subdifferential of the tensor ℓ_1-norm and a subset of the subdifferential of the TTNN, respectively.

Suppose that $\mathcal{A}, \mathcal{G} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, the subdifferential of any norm $|||\mathcal{A}|||$ can be given by

$$\partial |||\mathcal{A}||| = \left\{ \mathcal{G} : |||\mathcal{B}||| \geq |||\mathcal{A}||| + \text{Re}(\langle \mathcal{G}, \mathcal{B} - \mathcal{A} \rangle), \forall\, \mathcal{B} \in \mathbb{C}^{n_1 \times n_2 \times n_3} \right\}. \tag{B10}$$

Let $\tilde{\Omega}$ denote the locations that $\mathcal{A}_{ijk}$ are nonzeros.

Lemma 9. *Suppose that* $\mathcal{A}, \mathcal{G} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, *denote* $\partial \|\mathcal{A}\|_1$ *as the subdifferential of* $\|\cdot\|_1$ *at* $\mathcal{A}$ *which is supported on* $\tilde{\Omega}$, *then*

$$\partial \|\mathcal{A}\|_1 = \left\{ \mathcal{G} : \mathcal{G} = \texttt{dir}(\mathcal{A}) + \mathcal{F}, \mathcal{P}_{\tilde{\Omega}}(\mathcal{F}) = 0, \|\mathcal{F}\|_{\infty} \leq 1 \right\}, \tag{B11}$$

where $\texttt{dir}(\mathcal{A})$ *is defined as*

$$\texttt{dir}(\mathcal{A}_{ijk}) = \begin{cases} \dfrac{\mathcal{A}_{ijk}}{|\mathcal{A}_{ijk}|}, & \mathcal{A}_{ijk} \neq 0, \\ 0, & \mathcal{A}_{ijk} = 0. \end{cases}$$

Proof. For any $x \in \mathbb{C}$, the subdifferential of $|\cdot|$ is given by

$$\partial |x| = \begin{cases} x/|x|, & \text{if } x \neq 0, \\ [-1, 1], & \text{if } x = 0. \end{cases}$$

Since each entries of $\|\mathcal{A}\|_1$ is separable, we can obtain easily that

$$\partial \|\mathcal{A}\|_1 = \left\{ \mathcal{G} : \mathcal{G} = \texttt{dir}(\mathcal{A}) + \mathcal{F}, \mathcal{P}_{\tilde{\Omega}}(\mathcal{F}) = \mathbf{0}, \|\mathcal{F}\|_{\infty} \leq 1 \right\}.$$

This completes the proof. ∎

Lemma 10. *Suppose that $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ has the transformed tensor SVD as (B2). Denote $\partial \|\mathcal{A}\|_{\mathrm{TTNN}}$ as the subdifferential of $\|\cdot\|_{\mathrm{TTNN}}$ at $\mathcal{A}$, then*

$$\partial \|\mathcal{A}\|_{\mathrm{TTNN}} \supseteq \left\{ \mathcal{G} \in \mathbb{C}^{n_1 \times n_2 \times n_3} : \mathcal{G} = \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H + \mathcal{F} \right\}, \tag{B12}$$

where $\mathcal{U}^H \diamond_{\mathbf{\Phi}} \mathcal{F} = \mathcal{F}^H \diamond_{\mathbf{\Phi}} \mathcal{V} = \mathbf{0}$ and $\|\mathcal{F}\|_{\mathbf{\Phi}} \leq 1$.

Proof. If $\mathcal{A} = \mathbf{0}$, then the result is obvious. Now we assume $\mathcal{A} \neq \mathbf{0}$, and $\mathcal{Y}$ is given as the right hand of (B12). It follows that

$$\mathrm{Re}(\langle \mathcal{A}, \mathcal{Y} \rangle) = \mathrm{Re}\left(\langle \mathcal{A}, \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H + \mathcal{F} \rangle \right) = \mathrm{Re}\left(\langle \mathcal{A}, \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H \rangle \right) + \mathrm{Re}\left(\langle \mathcal{A}, \mathcal{F} \rangle \right) = \|\mathcal{A}\|_{\mathrm{TTNN}}.$$

Moreover,

$$\|\mathcal{Y}\|_{\mathbf{\Phi}} = \left\| \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H + \mathcal{F} \right\|_{\mathbf{\Phi}} = \max\left\{ \left\| \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H \right\|_{\mathbf{\Phi}} + \|\mathcal{F}\|_{\mathbf{\Phi}} \right\} = 1.$$

Then $\mathcal{Y}$ satisfies (B10), which shows that $\mathcal{Y}$ is a subgradient of the TTNN at $\mathcal{A}$. ∎

The argument of the proof of Theorem 2 can be divided into two steps. The first step is to show a sufficient condition for the pair $(\mathcal{L}_0, \mathcal{E}_0)$ to be the unique optimal solution to problem (5). The second step is to prove that when the required assumptions in Theorem 2 are satisfied, the sufficient conditions derived by the first step are satisfied. Now we give a sufficient condition for the pair $(\mathcal{L}_0, \mathcal{E}_0)$ to be the unique optimal solution to problem (5). The conditions are stated in terms of a dual variable $\mathcal{Y}$, which is given in Theorem 4.

Theorem 4. *Assume that there is a tensor $\mathcal{Y} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ obeying*

$$\begin{cases} \|\mathcal{P}_T(\mathcal{Y} + \lambda \mathtt{dir}(\mathcal{E}_0) - \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H)\|_F \leq \frac{\lambda}{n_1 n_2 n_3^2}, \\ \|\mathcal{P}_{T^\perp}(\mathcal{Y} + \lambda \mathtt{dir}(\mathcal{E}_0))\|_{\mathbf{\Phi}} \leq \frac{1}{2}, \\ \|\mathcal{P}_\Gamma(\mathcal{Y})\|_\infty \leq \frac{\lambda}{2}, \\ \mathcal{P}_{\Gamma^\perp}(\mathcal{Y}) = \mathbf{0}, \end{cases} \tag{B13}$$

where $\lambda = 1/\sqrt{\rho n_{(1)} n_3}$, then $(\mathcal{L}_0, \mathcal{E}_0)$ is the unique optimal solution to (5) when n_1, n_2, n_3 are sufficient large.

Proof. Let $f(\mathcal{L}, \mathcal{E}) := \|\mathcal{L}\|_{\mathrm{TTNN}} + \lambda \|\mathcal{E}\|_1$. Given a feasible perturbation $(\mathcal{L}_0 + \mathcal{Z}, \mathcal{E}_0 - \mathcal{P}_\Omega(\mathcal{Z}))$, we will show that the objective function value $f(\mathcal{L}_0 + \mathcal{Z}, \mathcal{E}_0 - \mathcal{P}_\Omega(\mathcal{Z}))$ is strictly larger than $f(\mathcal{L}_0, \mathcal{E}_0)$ unless $\mathcal{Z} = \mathbf{0}$. Denote the skinny transformed tensor SVD of $\mathcal{P}_{T^\perp}(\mathcal{Z})$ by $\mathcal{P}_{T^\perp}(\mathcal{Z}) = \mathcal{U}_\perp \diamond_{\mathbf{\Phi}} \mathcal{S}_\perp \diamond_{\mathbf{\Phi}} \mathcal{V}_\perp^H$. Then we have

$$\|\mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H + \mathcal{U}_\perp \diamond_{\mathbf{\Phi}} \mathcal{V}_\perp^H\|_{\mathbf{\Phi}} = 1.$$

It follows from Lemmas 9 and 10 that

$$\begin{aligned} \|\mathcal{L}_0 + \mathcal{Z}\|_{\mathrm{TTNN}} &\geq \mathrm{Re}(\langle \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H + \mathcal{U}_\perp \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \mathcal{L}_0 + \mathcal{Z} \rangle) \\ &= \mathrm{Re}(\langle \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \mathcal{L}_0 \rangle + \langle \mathcal{U}_\perp \diamond_{\mathbf{\Phi}} \mathcal{V}_\perp^H, \mathcal{P}_{T^\perp}(\mathcal{Z}) \rangle + \langle \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \mathcal{Z} \rangle) \\ &= \|\mathcal{L}_0\|_{\mathrm{TTNN}} + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}} + \mathrm{Re}(\langle \mathcal{U} \diamond_{\mathbf{\Phi}} \mathcal{V}^H, \mathcal{Z} \rangle), \end{aligned}$$

and

$$\begin{aligned} \|\mathcal{E}_0 - \mathcal{P}_\Omega(\mathcal{Z})\|_1 &= \|\mathcal{P}_\Omega(\mathcal{E}_0 - \mathcal{Z})\|_1 = \|\mathcal{P}_\Gamma(\mathcal{E}_0 - \mathcal{Z})\|_1 + \|\mathcal{P}_\Lambda(\mathcal{E}_0 - \mathcal{Z})\|_1 \\ &= \|\mathcal{P}_\Gamma(\mathcal{Z})\|_1 + \|\mathcal{E}_0 - \mathcal{P}_\Lambda(\mathcal{Z})\|_1 \geq \|\mathcal{P}_\Gamma(\mathcal{Z})\|_1 + \|\mathcal{E}_0\|_1 - \mathrm{Re}(\langle \mathtt{dir}(\mathcal{E}_0), \mathcal{P}_\Lambda(\mathcal{Z}) \rangle) \\ &\geq \|\mathcal{P}_\Gamma(\mathcal{Z})\|_1 + \|\mathcal{E}_0\|_1 - \mathrm{Re}(\langle \mathtt{dir}(\mathcal{E}_0), \mathcal{Z} \rangle). \end{aligned}$$

Therefore, we obtain that

$$
\begin{aligned}
&f(\mathcal{L}_0+\mathcal{Z},\mathcal{E}_0-\mathcal{P}_\Omega(\mathcal{Z}))-f(\mathcal{L}_0,\mathcal{E}_0)\\
&\quad=\|\mathcal{L}_0+\mathcal{Z}\|_{\mathrm{TTNN}}+\lambda\|\mathcal{E}_0-\mathcal{P}_\Omega(\mathcal{Z})\|_1-\|\mathcal{L}_0\|_{\mathrm{TTNN}}-\lambda\|\mathcal{E}_0\|_1\\
&\quad\geq\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}+\lambda\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1-\mathrm{Re}\langle\lambda\mathtt{dir}(\mathcal{E}_0)-\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H,\mathcal{Z}\rangle\\
&\quad\geq\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}+\lambda\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1-\mathrm{Re}\langle\mathcal{P}_\Gamma(\mathcal{Y}),\mathcal{P}_\Gamma(\mathcal{Z})\rangle\\
&\qquad-\mathrm{Re}\langle\mathcal{P}_T(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)-\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H),\mathcal{P}_T(\mathcal{Z})\rangle-\mathrm{Re}\langle\mathcal{P}_{T^\perp}(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)),\mathcal{P}_{T^\perp}(\mathcal{Z})\rangle\\
&\quad=\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}+\lambda\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1-\ \mathrm{Re}\langle\mathcal{P}_\Gamma(\mathcal{Y}),\mathcal{P}_\Gamma(\mathcal{Z})\rangle\\
&\qquad-\mathrm{Re}\langle\overline{(\mathcal{P}_T(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)-\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H))_{\mathbf{\Phi}}},\overline{(\mathcal{P}_T(\mathcal{Z}))_{\mathbf{\Phi}}}\rangle-\mathrm{Re}\langle\overline{(\mathcal{P}_{T^\perp}(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)))_{\mathbf{\Phi}}},\overline{(\mathcal{P}_{T^\perp}(\mathcal{Z}))_{\mathbf{\Phi}}}\rangle\\
&\quad\geq\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}+\lambda\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1-\|\mathcal{P}_\Gamma(\mathcal{Y})\|_\infty\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1\\
&\qquad-\|\overline{(\mathcal{P}_T(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)-\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H))_{\mathbf{\Phi}}}\|_F\|\overline{(\mathcal{P}_T(\mathcal{Z}))_{\mathbf{\Phi}}}\|_F-\|\overline{(\mathcal{P}_{T^\perp}(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)))_{\mathbf{\Phi}}}\|\|\overline{(\mathcal{P}_{T^\perp}(\mathcal{Z}))_{\mathbf{\Phi}}}\|_*\\
&\quad=\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}+\lambda\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1-\|\mathcal{P}_\Gamma(\mathcal{Y})\|_\infty\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1\\
&\qquad-\|\mathcal{P}_T(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0)-\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H)\|_F\|\mathcal{P}_T(\mathcal{Z})\|_F-\|\mathcal{P}_{T^\perp}(\mathcal{Y}+\lambda\mathtt{dir}(\mathcal{E}_0))\|_{\mathbf{\Phi}}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}\\
&\quad\geq\frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}+\frac{\lambda}{2}\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1-\frac{\lambda}{n_1n_2n_3^2}\|\mathcal{P}_T(\mathcal{Z})\|_F,
\end{aligned}
\tag{B14}
$$

where the inequality (B14) is due to (B13). It follows from Lemma 5 that

$$
\left\|\frac{1}{(1-2\gamma)\rho}\mathcal{P}_T\mathcal{P}_\Gamma\mathcal{P}_T-\mathcal{P}_T\right\|_{\mathrm{op}}\leq\frac{1}{2},
$$

which implies $\|\frac{1}{\sqrt{(1-2\gamma)\rho}}\mathcal{P}_T\mathcal{P}_\Gamma\|_{\mathrm{op}}\leq\sqrt{3/2}$. Therefore, we get

$$
\begin{aligned}
\|\mathcal{P}_T(\mathcal{Z})\|_F=\|\overline{(\mathcal{P}_T(\mathcal{Z}))_{\mathbf{\Phi}}}\|_F&\leq 2\left\|\frac{1}{(1-2\gamma)\rho}\overline{(\mathcal{P}_T\mathcal{P}_\Gamma\mathcal{P}_T(\mathcal{Z}))_{\mathbf{\Phi}}}\right\|_F\\
&\leq 2\left\|\frac{1}{(1-2\gamma)\rho}\overline{(\mathcal{P}_T\mathcal{P}_\Gamma\mathcal{P}_{T^\perp}(\mathcal{Z}))_{\mathbf{\Phi}}}\right\|_F+2\left\|\frac{1}{(1-2\gamma)\rho}\overline{(\mathcal{P}_T\mathcal{P}_\Gamma(\mathcal{Z}))_{\mathbf{\Phi}}}\right\|_F\\
&\leq\sqrt{\frac{6}{(1-2\gamma)\rho}}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F+\sqrt{\frac{6}{(1-2\gamma)\rho}}\|\mathcal{P}_\Gamma(\mathcal{Z})\|_F.
\end{aligned}
\tag{B15}
$$

It is easy to check that

$$
\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\mathrm{TTNN}}=\|\overline{(\mathcal{P}_{T^\perp}(\mathcal{Z}))_{\mathbf{\Phi}}}\|_*\geq\|\overline{(\mathcal{P}_{T^\perp}(\mathcal{Z}))_{\mathbf{\Phi}}}\|_F=\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F,
\tag{B16}
$$

and $\|\mathcal{P}_\Gamma(\mathcal{Z})\|_1\geq\|\mathcal{P}_\Gamma(\mathcal{Z})\|_F$. Plugging (B15) and (B16) into (B14), we have

$$
f(\mathcal{L}_0+\mathcal{Z},\mathcal{E}_0-\mathcal{P}_\Omega(\mathcal{Z}))-f(\mathcal{L}_0,\mathcal{E}_0)\geq\left(\frac{1}{2}-\frac{\lambda}{n_1n_2n_3^2}\sqrt{\frac{6}{(1-2\gamma)\rho}}\right)\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F+\left(\frac{\lambda}{2}-\frac{\lambda}{n_1n_2n_3^2}\sqrt{\frac{6}{(1-2\gamma)\rho}}\right)\|\mathcal{P}_\Gamma(\mathcal{Z})\|_F.
\tag{B17}
$$

When n_1,n_2,n_3 are sufficiently large such that

$$
\frac{1}{2}-\frac{\lambda}{n_1n_2n_3^2}\sqrt{\frac{6}{(1-2\gamma)\rho}}>0,\quad\frac{\lambda}{2}-\frac{\lambda}{n_1n_2n_3^2}\sqrt{\frac{6}{(1-2\gamma)\rho}}>0,
$$

the inequality (B17) holds if and only if $\mathcal{P}_{T^\perp}(\mathcal{Z})=\mathcal{P}_\Gamma(\mathcal{Z})=\mathbf{0}$. Thus, $\mathcal{P}_T\mathcal{P}_\Gamma(\mathcal{Z})=\mathbf{0}$. On the other hand, when ρ is sufficiently large and γ is sufficiently small (which are bounded by two constants c_ρ and c_γ), we obtain that

$$
\|\mathcal{P}_T\mathcal{P}_\Gamma\|_{\mathrm{op}}\leq\sqrt{\frac{3(1-2\gamma)\rho}{2}}<1,
$$

which implies that $\mathcal{P}_T\mathcal{P}_\Gamma$ is injective. As a result, (B17) holds if and only if $\mathcal{Z} = \mathbf{0}$. ∎

Similar to References 18 and 46, we use the golfing scheme to construct the dual tensor $\mathcal{Y}$, which is supported on Γ. Afterward, the size of Γ is increased gradually. Consider the set $\Gamma \sim \mathrm{Ber}((1-2\gamma)\rho)$ as a union of sets of support Γ_j, that is,

$$\Gamma = \bigcup_{j=1}^{p} \Gamma_j,$$

where $\Gamma_j \sim \mathrm{Ber}(q_j)$. Let $q_1 = q_2 = \frac{(1-2\gamma)\rho}{6}$ and $q_3 = \ldots = q_p = q$, which implies $q \geq C_0\rho/\log(n_{(1)}n_3)$. Hence we have

$$1-(1-2\gamma)\rho = \left(1-\frac{(1-2\gamma)\rho}{6}\right)^2(1-q)^{p-2},$$

where $p = \lfloor 5\log(n_{(1)}n_3)+1\rfloor$. Starting from $\mathcal{Z}_0 = \mathcal{P}_T(\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H - \lambda\mathrm{sgn}(\mathcal{E}_0))$, we define inductively

$$\mathcal{Z}_j = \left(\mathcal{P}_T - \frac{1}{q_j}\mathcal{P}_T\mathcal{P}_{\Gamma_j}\mathcal{P}_T\right)\mathcal{Z}_{j-1}.$$

Then it follows from Lemmas 5–7 that

$$\|\mathcal{Z}_j\|_F \leq \frac{1}{2}\|\mathcal{Z}_{j-1}\|_F, \quad j = 1,\ldots,p, \tag{B18}$$

$$\|\mathcal{Z}_1\|_\infty \leq \frac{1}{2\sqrt{\log(n_{(1)}n_3)}}\|\mathcal{Z}_0\|_\infty, \tag{B19}$$

$$\|\mathcal{Z}_j\|_\infty \leq \frac{1}{2^j\log(n_{(1)}n_3)}\|\mathcal{Z}_0\|_\infty, \quad j = 2,\ldots,p, \tag{B20}$$

and

$$\|(\mathcal{I}_{\mathbf{\Phi}} - q_j^{-1}\mathcal{P}_{\Gamma_j})\mathcal{Z}_{j-1}\|_{\mathbf{\Phi}} \leq C_0'\sqrt{\frac{n_{(1)}n_3\log(n_{(1)}n_3)}{q_j}}\|\mathcal{Z}_{j-1}\|_\infty, \quad j = 2,\ldots,p, \tag{B21}$$

with high probability provided c_r and c_γ are small enough.

Let the dual tensor $\mathcal{Y}$ be

$$\mathcal{Y} = \sum_{j=1}^{p}\frac{1}{q_j}\mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}). \tag{B22}$$

Then we need to show that the $\mathcal{Y}$ in (B22) satisfies (B13). Obviously, $\mathcal{P}_{\Gamma^\perp}(\mathcal{Y}) = \mathbf{0}$ and it is sufficient to prove

$$\begin{cases} \|\mathcal{P}_T(\mathcal{Y} + \lambda\mathtt{dir}(\mathcal{E}_0) - \mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H)\|_F \leq \frac{\lambda}{n_1n_2n_3^2}, \\ \|\mathcal{P}_{T^\perp}(\mathcal{Y})\|_{\mathbf{\Phi}} \leq \frac{1}{4}, \\ \lambda\|\mathcal{P}_{T^\perp}(\mathtt{dir}(\mathcal{E}_0))\|_{\mathbf{\Phi}} \leq \frac{1}{4}, \\ \|\mathcal{P}_\Gamma(\mathcal{Y})\|_\infty \leq \frac{\lambda}{2}, \end{cases} \tag{B23}$$

where $\lambda = 1/\sqrt{\rho n_{(1)}n_3}$ and n_1, n_2, n_3 are large enough.

First, let us bound $\|\mathcal{Z}_0\|_F$ and $\|\mathcal{Z}_0\|_\infty$. By the triangle inequality, we get

$$\|\mathcal{Z}_0\|_\infty \leq \|\mathcal{U}\diamond_{\mathbf{\Phi}}\mathcal{V}^H\|_\infty + \lambda\|\mathcal{P}_T(\mathtt{dir}(\mathcal{E}_0))\|_\infty. \tag{B24}$$

Note that $\mathcal{E}_{ijk} = \vec{\boldsymbol{e}}_{ik}\diamond_{\mathbf{\Phi}}\mathring{\boldsymbol{e}}_{kk}\diamond_{\mathbf{\Phi}}\vec{\boldsymbol{e}}_{jk}^H$, then $\mathtt{dir}(\mathcal{E}_0) = \sum_{i,j,k}(\mathtt{dir}(\mathcal{E}_0))_{i,j,k}\mathcal{E}_{ijk}$. Therefore, by (B4) we have

$$\mathcal{P}_T(\mathtt{dir}(\mathcal{E}_0)) = \sum_{i,j,k}(\mathtt{dir}(\mathcal{E}_0))_{i,j,k}\mathcal{P}_T(\mathcal{E}_{ijk}).$$

Hence, the (a,b,c)th entry of $\mathcal{P}_T(\texttt{dir}(\mathcal{E}_0))$ can be represented by

$$\langle \mathcal{P}_T(\texttt{dir}(\mathcal{E}_0)), \vec{\boldsymbol{e}}_{ac} \diamond_{\boldsymbol{\Phi}} \mathring{\boldsymbol{e}}_{cc} \diamond_{\boldsymbol{\Phi}} \vec{\boldsymbol{e}}_{bc}^{H} \rangle = \sum_{ijk} (\texttt{dir}(\mathcal{E}_0))_{i,j,k} \langle \mathcal{P}_T(\mathcal{E}_{ijk}), \mathcal{E}_{abc} \rangle.$$

By Bernstein's inequality, we further have

$$\mathbb{P}(|\langle \mathcal{P}_T(\texttt{dir}(\mathcal{E}_0)), \mathcal{E}_{abc} \rangle| \geq \tau) \leq 2 \exp\left(\frac{-\tau^2/2}{N + M\tau/3} \right),$$

where

$$M = \left| [\texttt{dir}(\mathcal{E}_0)]_{i,j,k} \right| \|\mathcal{P}_T(\mathcal{E}_{ijk})\|_F \|\mathcal{P}_T(\mathcal{E}_{abc})\|_F \leq \frac{2\mu r}{n_{(2)}},$$

and $N = 2\gamma\rho \|\mathcal{P}_T(\mathcal{E}_{ijk})\|_F^2 \leq 4\gamma\rho \frac{\mu r}{n_{(2)}}$. Since the entries of $\mathcal{P}_T(\texttt{dir}(\mathcal{E}_0))$ can be understood as i.i.d. copies of the (a,b,c)th entry, it follows from the union bound that

$$\|\mathcal{P}_T(\texttt{dir}(\mathcal{E}_0))\|_\infty \leq C'' \sqrt{\frac{\rho\mu r \log(n_{(1)} n_3)}{n_{(2)}}}, \tag{B25}$$

with high probability for some numerical constant C''. By the joint incoherence condition (8), we obtain that

$$\|\mathcal{U} \diamond_{\boldsymbol{\Phi}} \mathcal{V}^H\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2 n_3}} = \lambda \sqrt{\frac{\rho\mu r}{n_{(2)}}}. \tag{B26}$$

Therefore, combining (B24) and (B25), with (B26), we get

$$\|\mathcal{Z}_0\|_\infty \leq C\lambda \sqrt{\frac{\rho\mu r \log(n_{(1)} n_3)}{n_{(2)}}}, \tag{B27}$$

$$\|\mathcal{Z}_0\|_F \leq \sqrt{n_1 n_2 n_3} \|\mathcal{Z}_0\|_\infty \leq C\lambda \sqrt{\rho\mu r n_{(1)} n_3 \log(n_{(1)} n_3)}, \tag{B28}$$

where $C = \max\left\{ \frac{1}{\log(n_{(1)} n_3)}, C'' \right\}$.

Now, let us turn to the proof of (B23). By (B22), we obtain that

$$\begin{aligned}
\|\mathcal{P}_T(\mathcal{Y}) + \mathcal{P}_T(\lambda \texttt{dir}(\mathcal{E}_0) - \mathcal{U} \diamond_{\boldsymbol{\Phi}} \mathcal{V}^H)\|_F &= \left\| \mathcal{Z}_0 - \sum_{j=1}^{p} \frac{1}{q_j} \mathcal{P}_T \mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}) \right\|_F \\
&= \left\| \left(\mathcal{P}_T - \frac{1}{q_1} \mathcal{P}_T \mathcal{P}_{\Gamma_1} \mathcal{P}_T \right) \mathcal{Z}_0 - \sum_{j=2}^{p} \frac{1}{q_j} \mathcal{P}_T \mathcal{P}_{\Gamma_j} \mathcal{P}_T(\mathcal{Z}_{j-1}) \right\|_F \\
&= \left\| \mathcal{P}_T(\mathcal{Z}_1) - \sum_{j=2}^{p} \frac{1}{q_j} \mathcal{P}_T \mathcal{P}_{\Gamma_j} \mathcal{P}_T(\mathcal{Z}_{j-1}) \right\|_F = \ldots = \|\mathcal{Z}_p\|_F \\
&\leq \left(\frac{1}{2} \right)^p \|\mathcal{Z}_0\|_F \leq C \left(n_{(1)} n_3 \right)^{-5} \lambda \sqrt{\rho\mu r n_{(1)} n_3 \log(n_{(1)} n_3)} \\
&\leq \frac{\lambda}{n_1 n_2 n_3^2},
\end{aligned}$$

where the first inequality follows from (B18) and the second inequality from (B28).

Furthermore, provided that $\lambda = 1/\sqrt{\rho n_{(1)} n_3}$ and c_r is sufficiently small, we have

$$\begin{aligned}
\|\mathcal{P}_{T^\perp}(\mathcal{Y})\|_{\mathbf{\Phi}} &= \left\| \mathcal{P}_{T^\perp} \sum_{j=1}^{p} \frac{1}{q_j} \mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}) \right\|_{\mathbf{\Phi}} \\
&\le \sum_{j=1}^{p} \left\| \frac{1}{q_j} \mathcal{P}_{T^\perp} \mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}) \right\|_{\mathbf{\Phi}} = \sum_{j=1}^{p} \left\| \mathcal{P}_{T^\perp} \left(\frac{1}{q_j} \mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}) - \mathcal{Z}_{j-1} \right) \right\|_{\mathbf{\Phi}} \\
&\le \sum_{j=1}^{p} \left\| \frac{1}{q_j} \mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}) - \mathcal{Z}_{j-1} \right\|_{\mathbf{\Phi}} \le \sum_{j=1}^{p} C_0' \sqrt{\frac{n_{(1)} n_3 \log(n_{(1)} n_3)}{q_j}} \|\mathcal{Z}_{j-1}\|_\infty \\
&\le C_0' \sqrt{n_{(1)} n_3 \log(n_{(1)} n_3)} \left(\sum_{j=3}^{p} \frac{1}{2^{j-1} \log(n_{(1)} n_3) \sqrt{q_j}} + \frac{1}{2\sqrt{q_2 \log(n_{(1)} n_3)}} + \frac{1}{\sqrt{q_1}} \right) \|\mathcal{Z}_0\|_\infty \\
&\le C' \lambda \sqrt{\frac{\rho \mu r n_{(1)} n_3 (\log(n_{(1)} n_3))^2}{\rho n_{(2)}}} \\
&\le C' \sqrt{c_r} \le \frac{1}{4},
\end{aligned}$$

where the third inequality follows from Lemma 7, the fourth inequality follows from (B19) to (B21), and the fifth inequality follows from (B27), respectively.

Note that the direction tensor $\mathtt{dir}(\mathcal{E}_0)$ satisfies

$$\left| [\mathtt{dir}(\mathcal{E}_0)]_{i,j,k} \right| = \begin{cases} 1, & \text{with probability } \gamma\rho, \\ 0, & \text{with probability } 1 - \gamma\rho. \end{cases}$$

As proved by Lemma 8, there exists a function $\varphi(\gamma\rho)$ satisfying $\lim_{\gamma\rho \to 0^+} \varphi(\gamma\rho) = 0$, such that

$$\|\mathtt{dir}(\mathcal{E}_0)\|_{\mathbf{\Phi}} \le \varphi(\gamma\rho) \sqrt{n_{(1)} n_3},$$

with high probability, which yields

$$\lambda \|\mathcal{P}_{T^\perp}(\mathtt{dir}(\mathcal{E}_0))\|_{\mathbf{\Phi}} \le \lambda \|\mathtt{dir}(\mathcal{E}_0)\|_{\mathbf{\Phi}} \le \varphi(\gamma\rho)/\sqrt{\rho} \le \frac{1}{4},$$

as long as c_r and c_γ are sufficiently small.

Finally, when c_r is small enough, we can get

$$\begin{aligned}
\|\mathcal{P}_{\Gamma}(\mathcal{Y})\|_\infty &= \left\| \mathcal{P}_T \sum_{j=1}^{p} \frac{1}{q_j} \mathcal{P}_{\Gamma_j}(\mathcal{Z}_{j-1}) \right\|_\infty \le \sum_{j=1}^{p} \frac{1}{q_j} \|\mathcal{Z}_{j-1}\|_\infty \\
&\le \left(\sum_{j=3}^{p} \frac{1}{2^{j-1} \log(n_{(1)} n_3) \sqrt{q_j}} \right) \|\mathcal{Z}_0\|_\infty + \left(\frac{1}{2\sqrt{q_2 \log(n_{(1)} n_3)}} + \frac{1}{\sqrt{q_1}} \right) \|\mathcal{Z}_0\|_\infty \\
&\le C\lambda \sqrt{\frac{\mu r \log(n_{(1)} n_3)}{\rho n_{(2)}}} \le C\lambda \sqrt{\frac{c_r}{\log(n_{(1)} n_3)}} \le \frac{\lambda}{2},
\end{aligned}$$

where the first inequality follows from Lemma 6, the second inequality follows from (B19) to (B21), and the third inequality follows from (B27), respectively.