



# A Hybrid Norm for Guaranteed Tensor Recovery

Yihao Luo<sup>1</sup>, Andong Wang<sup>1,2\*</sup>, Guoxu Zhou<sup>1,3</sup> and Qibin Zhao<sup>2</sup>

<sup>1</sup>School of Automation, Guangdong University of Technology, Guangzhou, China, <sup>2</sup>RIKEN AIP, Tokyo, Japan, <sup>3</sup>Key Laboratory of Intelligent Detection and the Internet of Things in Manufacturing, Ministry of Education, Guangzhou, China

## OPEN ACCESS

### Edited by:

Peng Zhang,  
Tianjin University, China

### Reviewed by:

Jingyao Hou,  
Southwest University, China  
Yong Peng,  
Hangzhou Dianzi University, China  
Jing Lou,  
Changzhou Institute of Mechatronic  
Technology, China  
Guifu Lu,  
Anhui Polytechnic University, China

### \*Correspondence:

Andong Wang  
w.a.d@outlook.com

### Specialty section:

This article was submitted to  
Statistical and Computational Physics,  
a section of the journal  
Frontiers in Physics

Received: 28 February 2022

Accepted: 27 April 2022

Published: 13 July 2022

### Citation:

Luo Y, Wang A, Zhou G and Zhao Q  
(2022) A Hybrid Norm for Guaranteed  
Tensor Recovery.  
Front. Phys. 10:885402.  
doi: 10.3389/fphy.2022.885402

Benefiting from the superiority of tensor Singular Value Decomposition (t-SVD) in excavating low-rankness in the spectral domain over other tensor decompositions (like Tucker decomposition), t-SVD-based tensor learning has shown promising performance and become an emerging research topic in computer vision and machine learning very recently. However, focusing on modeling spectral low-rankness, the t-SVD-based models may be insufficient to exploit low-rankness in the original domain, leading to limited performance while learning from tensor data (like videos) that are low-rank in both original and spectral domains. To this point, we define a hybrid tensor norm dubbed the “Tubal + Tucker” Nuclear Norm (T2NN) as the sum of two tensor norms, respectively, induced by t-SVD and Tucker decomposition to simultaneously impose low-rankness in both spectral and original domains. We further utilize the new norm for tensor recovery from linear observations by formulating a penalized least squares estimator. The statistical performance of the proposed estimator is then analyzed by establishing upper bounds on the estimation error in both deterministic and non-asymptotic manners. We also develop an efficient algorithm within the framework of Alternating Direction Method of Multipliers (ADMM). Experimental results on both synthetic and real datasets show the effectiveness of the proposed model.

**Keywords:** tensor decomposition, tensor low-rankness, tensor SVD, tubal nuclear norm, tensor completion

## 1 INTRODUCTION

Thanks to the rapid progress of computer technology, data in tensor format (i.e., multi-dimensional array) are emerging in computer vision, machine learning, remote sensing, quantum physics, and many other fields, triggering an increasing need for tensor-based learning theory and algorithms [1–6]. In this paper, we carry out both theoretic and algorithmic research studies on tensor recovery from linear observations, which is a typical problem in tensor learning aiming to learn an unknown tensor when only a limited number of its noisy linear observations are available [7]. Tensor recovery finds applications in many industrial circumstances where the sensed or collected tensor data are polluted by unpredictable factors such as sensor failures, communication losses, occlusion by objects, shortage of instruments, and electromagnetic interferences [7–9], and is thus of both theoretical and empirical significance.

In general, reconstructing an unknown tensor from only a small number of its linear observations is hopeless, unless some assumptions on the underlying tensor are made [9]. The most commonly used assumption is that the underlying tensor possesses some kind of low-rankness which can significantly limit its degree of freedom, such that the signal can be estimated from a small but sufficient number of observations [7]. However, as a higher-order extension of matrix low-rankness,

the tensor low-rankness has many different characterizations due to the multiple definitions of tensor rank, e.g., the CANDECOMP/PARAFAC (CP) rank [10], Tucker rank [11], Tensor Train (TT) rank [12], and Tensor Ring (TR) rank [13]. As has been discussed in [7] from a signal processing standpoint, the above exemplified rank functions are defined in the original domain of the tensor signal and may thus be insufficient to model low-rankness in the spectral domain. The recently proposed tensor low-tubal-rankness [14] within the algebraic framework of tensor Singular Value Decomposition (t-SVD) [15] gives a kind of complement to it by exploiting low-rankness in the spectral domain defined via Discrete Fourier Transform (DFT), and has witnessed significant performance improvements in comparison with the original domain-based low-rankness for tensor recovery [6, 16, 17].

Despite the popularity of low-tubal-rankness, the fact that it is defined solely in the spectral domain also naturally poses a potential limitation on its usability to some tensor data that are low-rank in both spectral and original domains. To address this issue, we propose a hybrid tensor norm to encourage low-rankness in both spectral and original domains at the same time for tensor recovery in this paper. Specifically, the contributions of this work are four-fold:

- To simultaneously exploit low-rankness in both spectral and original domains, we define a new norm named T2NN as the sum of two tensor nuclear norms induced, respectively, by the t-SVD for spectral low-rankness and Tucker decomposition for original domain low-rankness.
- Then, we apply the proposed norm to tensor recovery by formulating a new tensor least squares estimator penalized by T2NN.
- Statistically, the statistical performance of the proposed estimator is analyzed by establishing upper bounds on the estimation error in both deterministic and non-asymptotic manners.
- Algorithmically, we propose an algorithm based on ADMM to compute the estimator and evaluate its effectiveness on three different types of real data.

The rest of this paper proceeds as follows. First, the notations and preliminaries of low-tubal-rankness and low-Tucker-rankness are introduced in **Section 2**. Then, we define the new norm and apply it to tensor recovery in **Section 3**. To understand the statistical behavior of the estimator, we establish an upper bound on the estimation error in **Section 4**. To compute the proposed estimator, we design an ADMM-based algorithm in **Section 5** with empirical performance reported in **Section 6**.

## 2 NOTATIONS AND PRELIMINARIES

**Notations.** We use lowercase boldface, uppercase boldface, and calligraphy letters to denote vectors (e.g.,  $\mathbf{v}$ ), matrices (e.g.,  $\mathbf{M}$ ), and tensors (e.g.,  $\mathcal{T}$ ), respectively. For any real numbers  $a, b$ , let  $a \vee b = \max\{a, b\}$  and  $a \wedge b = \min\{a, b\}$ . If the size of a tensor is not given explicitly, then it is in  $\mathbb{R}^{d_1 \times d_2 \times d_3}$ . We use  $c, c', c_1$ , etc., to

denote constants whose values can vary from line to line. For notational simplicity, let  $\bar{d} = (d_1 + d_2)d_3$  and  $d_{\setminus k} = d_1 d_2 d_3 / d_k$  for  $k = 1, 2, 3$ .

Given a matrix  $\mathbf{M} \in \mathbb{C}^{d_1 \times d_2}$ , its nuclear norm and spectral norm are defined as  $\|\mathbf{M}\|_* := \sum_i \sigma_i$  and  $\|\mathbf{M}\| := \max \sigma_i$ , respectively, where  $\{\sigma_i \mid i = 1, 2, \dots, d_1 \wedge d_2\}$  are its singular values. Given a tensor  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ , define its  $l_1$ -norm and F-norm as  $\|\mathcal{T}\|_1 := \|\text{vec}(\mathcal{T})\|_1$ ,  $\|\mathcal{T}\|_F := \|\text{vec}(\mathcal{T})\|_2$ , respectively, where  $\text{vec}(\cdot)$  denotes the vectorization operation of a tensor [18]. Given  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ , let  $\mathcal{T}^{(i)} := \mathcal{T}(:, :, i)$  denote its  $i$ th frontal slice. For any two (real or complex) tensors  $\mathcal{A}, \mathcal{B}$  of the same size, define their inner product as the inner product of their vectorizations  $\langle \mathcal{A}, \mathcal{B} \rangle := \langle \text{vec}(\mathcal{A}), \text{vec}(\mathcal{B}) \rangle$ . Other notations are introduced at their first appearance.

### 2.1 Spectral Rankness Modeled by t-SVD

The low-tubal-rankness defined within the algebraic framework of t-SVD is a typical example to characterize low-rankness in the spectral domain. We give some basic notions about t-SVD in this section.

**Definition 1 (t-product [15]).** Given  $\mathcal{T}_1 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  and  $\mathcal{T}_2 \in \mathbb{R}^{d_2 \times d_4 \times d_3}$ , their t-product  $\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2 \in \mathbb{R}^{d_1 \times d_4 \times d_3}$  is a tensor whose  $(i, j)$ -th tube  $\mathcal{T}(i, j, :) = \sum_{k=1}^{d_2} \mathcal{T}_1(i, k, :) \bullet \mathcal{T}_2(k, j, :)$ , where  $\bullet$  is the circular convolution [15].

**Definition 2 (tensor transpose [15]).** Let  $\mathcal{T}$  be a tensor of size  $d_1 \times d_2 \times d_3$ , then  $\mathcal{T}^T$  is the  $d_2 \times d_1 \times d_3$  tensor obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through  $d_3$ .

**Definition 3 (identity tensor [15]).** The identity tensor  $\mathcal{I} \in \mathbb{R}^{d \times d \times d_3}$  is a tensor whose first frontal slice is the  $d \times d$  identity matrix and all other frontal slices are zero.

**Definition 4 (f-diagonal tensor [15]).** A tensor is called f-diagonal if each frontal slice of the tensor is a diagonal matrix.

**Definition 5 (Orthogonal tensor [15]).** A tensor  $\mathcal{Q} \in \mathbb{R}^{d \times d \times d_3}$  is orthogonal if  $\mathcal{Q}^T * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^T = \mathcal{I}$ .

Then, t-SVD can be defined as follows.

**Definition 6 (t-SVD, tubal rank [15]).** Any tensor  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  has a tensor singular value decomposition as

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \quad (1)$$

where  $\mathcal{U} \in \mathbb{R}^{d_1 \times d_1 \times d_3}$ ,  $\mathcal{V} \in \mathbb{R}^{d_2 \times d_2 \times d_3}$  are orthogonal tensors and  $\mathcal{S} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  is an f-diagonal tensor. The tubal rank of  $\mathcal{T}$  is defined as the number of non-zero tubes of  $\mathcal{T}$ ,

$$\text{rank}_{\text{tb}}(\mathcal{T}) := \#\{i \mid \mathcal{S}(i, i, :) \neq \mathbf{0}\}, \quad (2)$$

where  $\#$  counts the number of elements in a set.

For convenience of analysis, the block diagonal matrix of 3-way tensors is also defined.

**Definition 7 (block-diagonal matrix [15]).** Let  $\bar{\mathcal{T}}$  denote the block-diagonal matrix of the tensor  $\tilde{\mathcal{T}}$  in the Fourier domain<sup>1</sup>, i.e.,

<sup>1</sup>The Fourier version  $\tilde{\mathcal{T}}$  is obtained by performing 1D-DFT on all tubes of  $\mathcal{T}$ , i.e.,  $\tilde{\mathcal{T}} = \text{fft}(\mathcal{T}, [1, 3]) \in \mathbb{C}^{d_1 \times d_2 \times d_3}$  in MATLAB.

$$\tilde{\mathbf{T}} := \begin{bmatrix} \tilde{\mathbf{T}}^{(1)} & & \\ & \ddots & \\ & & \tilde{\mathbf{T}}^{(d_3)} \end{bmatrix} \in \mathbb{C}^{d_1 d_3 \times d_2 d_3}. \quad (3)$$

Definition 8 (tubal nuclear norm, tensor spectral norm [17]). Given  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ , let  $\tilde{\mathcal{T}}$  be its Fourier version in  $\mathbb{C}^{d_1 \times d_2 \times d_3}$ . The Tubal Nuclear Norm (TNN)  $\|\cdot\|_{\text{tnn}}$  of  $\mathcal{T}$  is defined as the averaged nuclear norm of frontal slices of  $\tilde{\mathcal{T}}$ ,

$$\|\mathcal{T}\|_{\text{tnn}} := \frac{1}{d_3} \sum_{i=1}^{d_3} \|\tilde{\mathbf{T}}^{(i)}\|_*$$

whereas the tensor spectral norm  $\|\cdot\|$  is the largest spectral norm of the frontal slices,

$$\|\mathcal{T}\| := \max_{i \in [d_3]} \{\|\tilde{\mathbf{T}}^{(i)}\|\}.$$

We can see from Definition 8 that TNN captures low-rankness in the spectral domain and is thus more suitable for tensors with spectral low-rankness. As visual data (like images and videos) often process strong spectral low-rankness, it has achieved superior performance over many original domain-based nuclear norms in visual data restoration [6, 17].

## 2.2 Original Domain Low-Rankness Modeled by Tucker Decomposition

The low-Tucker-rankness is a classical higher-order extension of matrix low-rankness in the original domain and has been widely applied in computer vision and machine learning [19–21]. Given any  $K$ -way tensor  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ , its Tucker rank is defined as the following vector:

$$\vec{\mathbf{r}}_{\text{Tucker}}(\mathcal{T}) := (\text{rank}(\mathbf{T}_{(1)}), \dots, \text{rank}(\mathbf{T}_{(K)}))^T \in \mathbb{R}^K, \quad (4)$$

where  $\mathbf{T}_{(k)} \in \mathbb{R}^{d_k \times \prod_{i \neq k} d_i}$  denotes the mode- $k$  unfolding (matrix) of  $\mathcal{T}$  [18] obtained by concatenating all the mode- $k$  fibers of  $\mathcal{T}$  as column vectors. We can see that the Tucker rank measures the low-rankness of all the mode- $k$  unfoldings  $\mathbf{T}_{(k)}$  in the original domain.

Through relaxing the matrix rank in Eq. 4 to its convex envelope, i.e., the matrix nuclear norm, we get a convex relaxation of the Tucker rank, called Sum of Nuclear Norms (SNN) [20], which is defined as follows:

$$\|\mathcal{T}\|_{\text{snn}} := \sum_{k=1}^K \alpha_k \|\mathbf{T}_{(k)}\|_*, \quad (5)$$

where  $\alpha_k$ 's are positive constants satisfying  $\sum_k \alpha_k = 1$ . As a typical tensor low-rankness penalty in the original domain, SNN has found many applications in tensor recovery [19, 20, 22].

## 3 A HYBRID NORM FOR TENSOR RECOVERY

In this section, we first define a new norm to exploit low-rankness in both spectral and original domains and then use it to formulate a penalized tensor least squares estimator.

### 3.1 The Proposed Norm

Although TNN has shown superior performance in many tensor learning tasks, it may still be insufficient for tensors which are low-rank in both spectral and original domains due to its definition solely in the spectral domain. Moreover, it is also unsuitable for tensors which have less significant spectral low-rankness than the original domain low-rankness. Thus, it is necessary to extend the vanilla TNN such that the original domain low-rankness can also be exploited for sounder low-rank modeling.

Under the inspiration of SNN's impressive low-rank modeling capability in the original domain, our idea is quite simple: to combine the advantages of both TNN and SNN through their weighted sum. In this line of thinking, we come up with the following hybrid tensor norm.

Definition 9 (T2NN). The hybrid norm called "Tubal + Tucker" Nuclear Norm (T2NN) of any 3-way tensor  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  is defined as the weighted sum of its TNN and SNN as follows:

$$\|\mathcal{T}\|_{\text{t2nn}} := \gamma \|\mathcal{T}\|_{\text{tnn}} + (1 - \gamma) \|\mathcal{T}\|_{\text{snn}}, \quad (6)$$

where  $\gamma \in (0, 1)$  is a constant balancing the low-rank modeling in the spectral and original domains.

As can be seen from its definition, T2NN approximates TNN when  $\gamma \rightarrow 1$ , and it degenerates to SNN as  $\gamma \rightarrow 0$ . Thus, it can be viewed as an interpolation between TNN and SNN, which provides with more flexibility in low-rank tensor modeling. We also define the dual norm of T2NN (named the dual T2NN norm) which are frequently used in analyzing the statistical performance of the T2NN-based tensor estimator.

Lemma 1. The dual norm of the proposed T2NN defined as

$$\|\mathcal{T}\|_{\text{t2nn}}^* := \sup_{\mathcal{X}} \langle \mathcal{X}, \mathcal{T} \rangle, \text{ s.t. } \|\mathcal{X}\|_{\text{t2nn}} \leq 1, \quad (7)$$

can be equivalently formulated as follows:

$$\|\mathcal{T}\|_{\text{t2nn}}^* = \inf_{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}} \max \left\{ \frac{1}{\gamma} \|\mathcal{A}\|_*, \frac{1}{\alpha_1(1-\gamma)} \|\mathbf{B}_{(1)}\|, \frac{1}{\alpha_2(1-\gamma)} \|\mathbf{C}_{(2)}\|, \frac{1}{\alpha_3(1-\gamma)} \|\mathbf{D}_{(3)}\| \right\}, \quad (8)$$

s.t.  $\mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} = \mathcal{T}$ .

Proof of Lemma 1. Using the definition of T2NN, the supremum in Problem (7) can be equivalently converted to the opposite number of infimum as follows:

$$-\|\mathcal{T}\|_{\text{t2nn}}^* = \inf_{\mathcal{X}} -\langle \mathcal{X}, \mathcal{T} \rangle, \quad (9)$$

s.t.  $\gamma \|\mathcal{X}\|_{\text{tnn}} + \alpha_1(1-\gamma) \|\mathcal{X}_{(1)}\|_* + \alpha_2(1-\gamma) \|\mathcal{X}_{(2)}\|_* + \alpha_3(1-\gamma) \|\mathcal{X}_{(3)}\|_* \leq 1$ .

By introducing a multiplier  $\lambda \geq 0$ , we obtain the Lagrangian function of Problem (9),

$$L(\mathcal{X}, \lambda) := -\langle \mathcal{X}, \mathcal{T} \rangle + \lambda (\gamma \|\mathcal{X}\|_{\text{tnn}} + \alpha_1(1-\gamma) \|\mathcal{X}_{(1)}\|_* + \alpha_2(1-\gamma) \|\mathcal{X}_{(2)}\|_* + \alpha_3(1-\gamma) \|\mathcal{X}_{(3)}\|_* - 1).$$

Since Slater's condition [23] is satisfied in Problem (9), strong duality holds, which means

$$-\|\mathcal{T}\|_{\text{t2nn}} = \inf_{\mathcal{X}} \sup_{\lambda} L(\mathcal{X}, \lambda) = \sup_{\lambda} \inf_{\mathcal{X}} L(\mathcal{X}, \lambda).$$

Thus, we proceed by computing  $\sup_{\lambda} \inf_{\mathcal{X}} L(\mathcal{X}, \lambda)$  as follows:

$$\begin{aligned} & \sup_{\lambda} \inf_{\mathcal{X}} -\langle \mathcal{X}, \mathcal{T} \rangle + \lambda(\gamma \|\mathcal{X}\|_{\text{tnn}} + \alpha_1(1-\gamma) \|\mathcal{X}_{(1)}\|_* \\ & \quad + \alpha_2(1-\gamma) \|\mathcal{X}_{(2)}\|_* + \alpha_3(1-\gamma) \|\mathcal{X}_{(3)}\|_* - 1) \\ & \stackrel{(i)}{=} \sup_{\lambda} \inf_{\mathcal{X}} -\langle \mathcal{X}, \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} \rangle \\ & \quad + \lambda(\gamma \|\mathcal{X}\|_{\text{tnn}} + \alpha_1(1-\gamma) \|\mathcal{X}_{(1)}\|_* + \alpha_2(1-\gamma) \|\mathcal{X}_{(2)}\|_* \\ & \quad + \alpha_3(1-\gamma) \|\mathcal{X}_{(3)}\|_* - 1) - \lambda, \quad \text{where} \\ & \quad \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} = \mathcal{T}, \\ & = \sup_{\lambda} \inf_{\mathcal{X}} -\lambda + (\lambda\gamma \|\mathcal{X}\|_{\text{tnn}} - \langle \mathcal{X}, \mathcal{A} \rangle) + (\lambda\alpha_1(1-\gamma) \\ & \quad \times \|\mathbf{X}_{(1)}\|_* - \langle \mathcal{X}, \mathcal{B} \rangle) + (\lambda\alpha_2(1-\gamma) \|\mathbf{X}_{(2)}\|_* - \langle \mathcal{X}, \mathcal{C} \rangle) \\ & \quad + (\lambda\alpha_3(1-\gamma) \|\mathbf{X}_{(3)}\|_* - \langle \mathcal{X}, \mathcal{D} \rangle), \quad \text{where} \\ & \quad \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} = \mathcal{T}, \\ & \stackrel{(ii)}{=} \sup_{\lambda} -\lambda + \begin{cases} 0 & \text{if } \lambda \geq \frac{1}{\gamma} \|\mathcal{A}\| \\ -\infty & \text{otherwise} \end{cases} \\ & \quad + \begin{cases} 0 & \text{if } \lambda \geq \frac{1}{\alpha_1(1-\gamma)} \|\mathcal{B}_{(1)}\| \\ -\infty & \text{otherwise} \end{cases} \\ & \quad + \begin{cases} 0 & \text{if } \lambda \geq \frac{1}{\alpha_2(1-\gamma)} \|\mathcal{C}_{(2)}\| \\ -\infty & \text{otherwise} \end{cases} \\ & \quad + \begin{cases} 0 & \text{if } \lambda \geq \frac{1}{\alpha_3(1-\gamma)} \|\mathcal{D}_{(3)}\| \\ -\infty & \text{otherwise} \end{cases} \\ & \quad \text{where } \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} = \mathcal{T}, \\ & = - \inf_{\mathcal{A}+\mathcal{B}+\mathcal{C}+\mathcal{D}=\mathcal{T}} \max \left\{ \frac{1}{\gamma} \|\mathcal{A}\|, \frac{1}{\alpha_1(1-\gamma)} \right. \\ & \quad \left. \times \|\mathcal{B}_{(1)}\|, \frac{1}{\alpha_2(1-\gamma)} \|\mathcal{C}_{(2)}\|, \frac{1}{\alpha_3(1-\gamma)} \|\mathcal{D}_{(3)}\| \right\}, \end{aligned}$$

where (i) is obtained by the trick of splitting  $\mathcal{T}$  into four auxiliary tensors  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$  for simpler analysis and (ii) holds because for any positive constant  $\alpha$ , any norm  $f(\cdot)$  with dual norm  $f^*(\cdot)$ , we have the following relationship:

$$\begin{aligned} \inf_{\mathcal{X}} \lambda \alpha f(\mathcal{X}) - \langle \mathcal{X}, \mathcal{A} \rangle & \geq \inf_{\mathcal{X}} \lambda \alpha f(\mathcal{X}) - \gamma f(\mathcal{X}) \cdot \frac{1}{\alpha} f^*(\mathcal{A}), \\ & = \inf_{\mathcal{X}} \alpha f(\mathcal{X}) \left( \lambda - \frac{1}{\alpha} f^*(\mathcal{A}) \right), \\ & = \begin{cases} 0, & \text{if } \lambda \geq \frac{1}{\alpha} f^*(\mathcal{A}), \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

This completes the proof.

Although an expression of the dual T2NN norm is given in Lemma 1, it is still an optimization problem whose optimal value cannot be straightforwardly computed from the variable tensor  $\mathcal{T}$ . Following the

tricks in [22], we instead give an upper bound on the dual T2NN norm which is directly in terms of  $\mathcal{T}$  in the following lemma:

**Lemma 2.** *The dual T2NN norm can be upper bounded as follows:*

$$\|\mathcal{T}\|_{\text{t2nn}}^* \leq \frac{1}{16} \left( \frac{1}{\gamma} \|\mathcal{T}\| + \frac{1}{\alpha_1(1-\gamma)} \|\mathbf{T}_{(1)}\| + \frac{1}{\alpha_2(1-\gamma)} \|\mathbf{T}_{(2)}\| + \frac{1}{\alpha_3(1-\gamma)} \|\mathbf{T}_{(3)}\| \right). \quad (10)$$

*Proof of Lemma 2.* The proof is a direct application of the basic equality “harmonic mean  $\leq$  arithmetic mean” with careful construction of auxiliary tensors  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$  in Eq. 8 as follows:

$$\begin{aligned} \mathcal{A}_0 & = \frac{\gamma \|\mathcal{T}\|^{-1}}{M}, \mathcal{B}_0 = \frac{\alpha_1(1-\gamma) \|\mathbf{T}_{(1)}\|^{-1}}{M}, \\ \mathcal{C}_0 & = \frac{\alpha_2(1-\gamma) \|\mathbf{T}_{(2)}\|^{-1}}{M}, \mathcal{D}_0 = \frac{\alpha_3(1-\gamma) \|\mathbf{T}_{(3)}\|^{-1}}{M}, \end{aligned}$$

where the denominator  $M$  is given by

$$M = \gamma \|\mathcal{T}\|^{-1} + \alpha_1(1-\gamma) \|\mathbf{T}_{(1)}\|^{-1} + \alpha_2(1-\gamma) \|\mathbf{T}_{(2)}\|^{-1} + \alpha_3(1-\gamma) \|\mathbf{T}_{(3)}\|^{-1}.$$

It is obvious that  $\mathcal{A}_0 + \mathcal{B}_0 + \mathcal{C}_0 + \mathcal{D}_0 = \mathcal{T}$ . By substituting the particular setting  $(\mathcal{A}_0, \mathcal{B}_0, \mathcal{C}_0, \mathcal{D}_0)$  of  $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$  into Eq. 8, we obtain

$$\|\mathcal{T}\|_{\text{t2nn}}^* \leq \frac{1}{\gamma \|\mathcal{T}\|^{-1} + \alpha_1(1-\gamma) \|\mathbf{T}_{(1)}\|^{-1} + \alpha_2(1-\gamma) \|\mathbf{T}_{(2)}\|^{-1} + \alpha_3(1-\gamma) \|\mathbf{T}_{(3)}\|^{-1}}. \quad (11)$$

Then, by using “harmonic mean  $\leq$  arithmetic mean” on the right-hand side of Eq. 11, we obtain

$$\begin{aligned} & \frac{4}{\gamma \|\mathcal{T}\|^{-1} + \alpha_1(1-\gamma) \|\mathbf{T}_{(1)}\|^{-1} + \alpha_2(1-\gamma) \|\mathbf{T}_{(2)}\|^{-1} + \alpha_3(1-\gamma) \|\mathbf{T}_{(3)}\|^{-1}} \\ & \leq \frac{1}{4} \left( \frac{1}{\gamma} \|\mathcal{T}\| + \frac{1}{\alpha_1(1-\gamma)} \|\mathbf{T}_{(1)}\| + \frac{1}{\alpha_2(1-\gamma)} \|\mathbf{T}_{(2)}\| + \frac{1}{\alpha_3(1-\gamma)} \|\mathbf{T}_{(3)}\| \right), \end{aligned} \quad (12)$$

which directly leads to Eq. 10.

## 3.2 T2NN-Based Tensor Recovery

### 3.2.1 The observation Model

We use  $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  to denote the underlying tensor which is unknown. Suppose one observes  $N \ll d_1 d_2 d_3$  scalars,

$$y_i = \langle \mathcal{L}^*, \mathcal{X}_i \rangle + \sigma \xi_i, \quad \forall i \in [N], \quad (13)$$

where  $\mathcal{X}_i$ 's are known (deterministic or random) design tensors,  $\xi_i$ 's are *i. i. d.* standard Gaussian noises, and  $\sigma$  is a known standard deviation constant measuring the noise level.

Let  $\mathbf{y} = (y_1, \dots, y_N)^\top$  and  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)^\top$  denote the collection of observations and noises. Define the design operator  $\mathfrak{X}(\cdot)$  with adjoint operator  $\mathfrak{X}^*(\cdot)$  as follows:

$$\begin{aligned} \forall \mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}, \quad \mathfrak{X}(\mathcal{T}) & := (\langle \mathcal{T}, \mathcal{X}_1 \rangle, \dots, \langle \mathcal{T}, \mathcal{X}_N \rangle)^\top \in \mathbb{R}^N, \\ \forall \mathbf{z} \in \mathbb{R}^N, \quad \mathfrak{X}^*(\mathbf{z}) & := \sum_{i=1}^N z_i \mathcal{X}_i \in \mathbb{R}^{d_1 \times d_2 \times d_3}. \end{aligned} \quad (14)$$

Then, the observation model (13) can be rewritten in the following compact form:

$$\mathbf{y} = \mathfrak{X}(\mathcal{L}^*) + \xi.$$

### 3.2.2 Two Typical Settings

With different settings of the design tensors  $\{\mathcal{X}_i\}$ , we consider two classical examples in this paper:

- *Tensor completion.* In tensor completion, the design tensors  $\{\mathcal{X}_i\}$  are *i. i. d.* random tensor bases drawn from uniform distribution on the canonical basis in the space of  $d_1 \times d_2 \times d_3$  tensors  $\{\mathbf{e}_i \circ \mathbf{e}_j \circ \mathbf{e}_k, \forall (i, j, k) \in [d_1] \times [d_2] \times [d_3]\}$ , where  $\mathbf{e}_i$  denotes the vector whose  $i$ th entry is 1 with all the other entries 0 and  $\circ$  denotes the tensor outer product [18].
- *Tensor compressive sensing.* When  $\mathfrak{X}$  is a *random Gaussian design*, Model (13) is the tensor compressive sensing model with Gaussian measurements [24].  $\mathfrak{X}$  is named a random Gaussian design when  $\{\mathcal{X}_i\}$  are random tensors with *i. i. d.* standard Gaussian entries [22].

### 3.2.3 The Proposed Estimator

The goal of this paper is to recover the unknown low-rank tensor  $\mathcal{L}^*$  from noisy linear observations  $\mathbf{y}$  satisfying the observation model (13).

Inspired by the capability of the newly defined T2NN in simultaneously modeling low-rankness in both spectral and original domains, we define the T2NN penalized least squares estimator  $\hat{\mathcal{L}}$  to estimate the unknown truth  $\mathcal{L}^*$ ,

$$\hat{\mathcal{L}} \in \operatorname{argmin}_{\mathcal{L}} \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2^2 + \lambda \|\mathcal{L}\|_{\text{t2nn}}, \quad (15)$$

where the squared  $l_2$ -norm is adopted as the fidelity term for Gaussian noises, the proposed T2NN is used to impose both spectral and original low-rankness in the solution, and  $\lambda$  is a penalization parameter which balances the residual fitting accuracy and the parameter complicity (characterized by low-rankness) of the model.

Given the estimator  $\hat{\mathcal{L}}$  in Eq. 15, one may naturally ask how well it can estimate the truth  $\mathcal{L}^*$  and how to compute it. In the following two sections, we first study the estimation performance of  $\hat{\mathcal{L}}$  by upper bounding its estimation error and then develop an ADMM-based algorithm to efficiently compute it.

## 4 STATISTICAL GUARANTEE

In this section, we first come up with a deterministic upper bound of the estimation error and then establish non-asymptotic error bounds for the special cases of tensor compressive sensing with random Gaussian design and noisy tensor completion.

First, to describe the low-rankness of  $\mathcal{L}^*$ , we consider both its low-tubal-rank and low-Tucker-rank structures as follows:

- *Low-tubal-rank structure:* Let  $r_{\text{tb}}^*$  denote the tubal rank of  $\mathcal{L}^*$ . Suppose it has reduced t-SVD  $\mathcal{L}^* = \mathbf{U} * \mathcal{S} * \mathbf{V}^\top$ , where  $\mathbf{U} \in \mathbb{R}^{d_1 \times r_{\text{tb}}^* \times d_3}$ ,  $\mathbf{V}^\top \in \mathbb{R}^{d_2 \times r_{\text{tb}}^* \times d_3}$  are orthogonal tensors and  $\mathcal{S} \in \mathbb{R}^{r_{\text{tb}}^* \times r_{\text{tb}}^* \times d_3}$  is f-diagonal. Then, following [25], we define the following projections of any tensor  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ :

$$P^\perp(\mathcal{T}) := (\mathcal{T} - \mathbf{U} * \mathbf{U}^\top) * \mathcal{T} * (\mathcal{T} - \mathbf{V} * \mathbf{V}^\top) \quad \text{and} \\ P(\mathcal{T}) = \mathcal{T} - P^\perp(\mathcal{T}) \quad (16)$$

where  $\mathcal{T}$  denotes the identity tensor of appropriate dimensionality.

- *Low-Tucker-rank structure:* Let  $\mathbf{r}_{\text{tk}}^* = (r_1^*, r_2^*, r_3^*)^\top$  denote the Tucker rank of  $\mathcal{L}^*$ , i.e.,  $\mathbf{r}_k^* = \mathbf{rank}(\mathbf{L}_{(k)}^*)$ . Then, we have the reduced SVD factorization  $\mathbf{T}_{(k)} = \mathbf{U}_k \mathcal{S}_k \mathbf{V}_k^\top$ , where  $\mathbf{U}_k \in \mathbb{R}^{d_k \times r_k^*}$  and  $\mathbf{V}_k \in \mathbb{R}^{(d_{\setminus k}) \times r_k^*}$  are orthogonal and  $\mathcal{S}_k \in \mathbb{R}^{r_k^* \times r_k^*}$  is diagonal. Let  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  be an arbitrary tensor. Similar to [22], we define the following two projections for any mode  $k = 1, 2, 3$ :

$$P_k^\perp(\mathcal{T}) = (\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^\top) \mathbf{T}_{(k)} (\mathbf{I} - \mathbf{V}_k \mathbf{V}_k^\top) \quad \text{and} \\ P_k(\mathcal{T}) = \mathbf{T}_{(k)} - P_k^\perp(\mathcal{T}), \quad (17)$$

where  $\mathbf{I}$  denotes the identity matrix of appropriate dimensionality.

## 4.1 A Deterministic Bound on the Estimation Error

Before bounding the Frobenius-norm error  $\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}$ , the particularity of the error tensor  $\Delta := \hat{\mathcal{L}} - \mathcal{L}^*$  is first characterized by a certain choice of regularization parameter  $\lambda$  involving the dual T2NN norm in the following proposition.

**Proposition 1.** *By setting the regularization parameter  $\lambda \geq 2\sigma \|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$ , we have*

(I) *rank inequality:*

$$\mathbf{rank}_{\text{tb}}(P(\Delta)) \leq 2r_{\text{tb}}^*, \quad \text{and} \\ \mathbf{rank}(P_k(\Delta)) \leq 2r_k^*, \quad k = 1, 2, 3, \quad (18)$$

(II) *sum of norms inequality:*

$$\gamma \|P^\perp(\Delta)\|_{\text{tnn}} + (1 - \gamma) \sum_{k=1}^3 \alpha_k \|P_k^\perp(\Delta)\|_* \\ \leq 3 \left( \gamma \|P(\Delta)\|_{\text{tnn}} + (1 - \gamma) \sum_{k=1}^3 \alpha_k \|P_k(\Delta)\|_* \right), \quad (19)$$

(III) *an upper bound on the “observed” error:*

$$\|\mathfrak{X}(\Delta)\|_2^2 \leq 3 \left( \gamma \sqrt{2r_{\text{tb}}^*} + (1 - \gamma) \sum_{k=1}^3 \alpha_k \sqrt{2r_k^*} \right) \|\Delta\|_{\text{F}}. \quad (20)$$

*Proof of Proposition 1.* The proof is given as follows:

*Proof of Part (I):* According to the definition of  $P(\mathcal{T})$  in Eq. 16, we have

$$\begin{aligned} P(\mathcal{T}) &= \mathcal{T} - P^\perp(\mathcal{T}) = \mathbf{U} * \mathbf{U}^\top * \mathcal{T} + \mathcal{T} * \mathbf{V} * \mathbf{V}^\top \\ &\quad - \mathbf{U} * \mathbf{U}^\top * \mathcal{T} * \mathbf{V} * \mathbf{V}^\top, \\ &= \mathbf{U} * \mathbf{U}^\top * \mathcal{T} + (\mathcal{I} - \mathbf{U} * \mathbf{U}^\top) \\ &\quad * \mathcal{T} * \mathbf{V} * \mathbf{V}^\top. \end{aligned}$$

Due to the facts that  $\text{rank}_{\text{tb}}(\mathcal{A} * \mathcal{B}) \leq \max\{\text{rank}_{\text{tb}}(\mathcal{A}), \text{rank}_{\text{tb}}(\mathcal{B})\}$ ,  $\text{rank}_{\text{tb}}(\mathcal{A} + \mathcal{B}) \leq \text{rank}_{\text{tb}}(\mathcal{A}) + \text{rank}_{\text{tb}}(\mathcal{B})$  [26], and  $\text{rank}_{\text{tb}}(\mathbf{U}) = \text{rank}_{\text{tb}}(\mathbf{V}) = r_{\text{tb}}^*$ , we have

$$\begin{aligned} \text{rank}_{\text{tb}}(P(\mathcal{T})) &\leq \text{rank}_{\text{tb}}(\mathbf{U} * \mathbf{U}^\top * \mathcal{T}) \\ &\quad + \text{rank}_{\text{tb}}((\mathcal{I} - \mathbf{U} * \mathbf{U}^\top) * \mathcal{T} * \mathbf{V} * \mathbf{V}^\top) \leq 2r_{\text{tb}}^*. \end{aligned}$$

Also, according to the definition of  $P(\mathcal{T})$  in Eq. 17, we have

$$\begin{aligned} P_k(\mathcal{T}) &= \mathbf{T}_{(k)} - P_k^\perp(\mathcal{T}) = \mathbf{U}_k \mathbf{U}_k^\top \mathbf{T}_{(k)} + \mathbf{T}_{(k)} \mathbf{V}_k \mathbf{V}_k^\top \\ &\quad - \mathbf{U}_k \mathbf{U}_k^\top \mathbf{T}_{(k)} \mathbf{V}_k \mathbf{V}_k^\top \\ &= \mathbf{U}_k \mathbf{U}_k^\top \mathbf{T}_{(k)} + (\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^\top) \\ &\quad \times \mathbf{T}_{(k)} \mathbf{V}_k \mathbf{V}_k^\top. \end{aligned}$$

Due to the facts that  $\text{rank}(\mathbf{A}\mathbf{B}) \leq \max\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$ ,  $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$  [26], and  $\text{rank}(\mathbf{U}_k) = \text{rank}(\mathbf{V}_k) = r_k^*$ , we have

$$\begin{aligned} \text{rank}(P_k(\mathcal{T})) &\leq \text{rank}(\mathbf{U}_k \mathbf{U}_k^\top * \mathbf{T}_{(k)}) \\ &\quad + \text{rank}((\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^\top) \mathbf{T}_{(k)} \mathbf{V}_k \mathbf{V}_k^\top) \leq 2r_k^*. \end{aligned}$$

*Proof of Part (II) and Part (III):* The optimality of  $\hat{\mathcal{L}}$  to Problem Eq. 15 indicates

$$\frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\hat{\mathcal{L}})\|_2^2 + \lambda \|\hat{\mathcal{L}}\|_{\text{t2nn}} \leq \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L}^*)\|_2^2 + \lambda \|\mathcal{L}^*\|_{\text{t2nn}}.$$

By the definition of the error tensor  $\Delta := \hat{\mathcal{L}} - \mathcal{L}^*$ , we can get  $\mathfrak{X}(\hat{\mathcal{L}}) = \mathfrak{X}(\mathcal{L}^*) + \mathfrak{X}(\Delta)$ , which leads to

$$\begin{aligned} \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L}^*) - \mathfrak{X}(\Delta)\|_2^2 - \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L}^*)\|_2^2 \\ \leq \lambda \|\mathcal{L}^*\|_{\text{t2nn}} - \lambda \|\hat{\mathcal{L}}\|_{\text{t2nn}}. \end{aligned}$$

The definition that  $\sigma\xi = \mathbf{y} - \mathfrak{X}(\mathcal{L}^*)$  yields

$$\begin{aligned} \frac{1}{2} \|\mathfrak{X}(\Delta)\|_2^2 &\leq \langle \mathfrak{X}(\Delta), \sigma\xi \rangle + \lambda (\|\mathcal{L}^*\|_{\text{t2nn}} - \|\hat{\mathcal{L}}\|_{\text{t2nn}}) \\ &\leq \langle \mathfrak{X}^*(\xi), \Delta \rangle + \lambda (\|\mathcal{L}^*\|_{\text{t2nn}} - \|\hat{\mathcal{L}}\|_{\text{t2nn}}), \end{aligned}$$

where the last inequality holds due to the definition of the adjoint operator  $\mathfrak{X}^*(\cdot)$ .

According to the definition and upper bound of the dual T2NN norm in Lemma 1 and Lemma 2, we obtain

$$\frac{1}{2} \|\mathfrak{X}(\Delta)\|_2^2 \leq \sigma \|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^* \|\Delta\|_{\text{t2nn}} + \lambda (\|\mathcal{L}^*\|_{\text{t2nn}}^* - \|\hat{\mathcal{L}}\|_{\text{t2nn}}). \quad (21)$$

According to the decomposibility of TNN (see the supplementary material of [25]) and the decomposibility of matrix nuclear norm [27], one has

$$\begin{aligned} \|\mathcal{L}^*\|_{\text{tnn}} - \|\hat{\mathcal{L}}\|_{\text{tnn}} &= \|\mathcal{L}^*\|_{\text{tnn}} - \|\mathcal{L}^* + \Delta\|_{\text{tnn}} \\ &= \|\mathcal{L}^*\|_{\text{tnn}} - \|\mathcal{L}^* + P^\perp(\Delta) + P(\Delta)\| \\ &\leq \|\mathcal{L}^*\|_{\text{tnn}} - (\|\mathcal{L}^* + P^\perp(\Delta)\|_{\text{tnn}} - \|P(\Delta)\|_{\text{tnn}}) \\ &= \|\mathcal{L}^*\|_{\text{tnn}} - (\|\mathcal{L}^*\|_{\text{tnn}} + \|P^\perp(\Delta)\|_{\text{tnn}} - \|P(\Delta)\|_{\text{tnn}}) \\ &= \|P(\Delta)\|_{\text{tnn}} - \|P^\perp(\Delta)\|_{\text{tnn}} \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{L}_{(k)}^*\|_* - \|\hat{\mathbf{L}}_{(k)}\|_* &= \|\mathbf{L}_{(k)}^*\|_* - \|\mathbf{L}_{(k)}^* + \hat{\Delta}_{(k)}\|_* \\ &= \|\mathbf{L}_{(k)}^*\|_* - \|(\mathbf{L}_{(k)}^* + P_k^\perp(\Delta)) + P_k(\Delta)\|_* \\ &\leq \|\mathbf{L}_{(k)}^*\|_* - (\|\mathbf{L}_{(k)}^* + P_k^\perp(\Delta)\|_* - \|P_k(\Delta)\|_*) \\ &= \|\mathbf{L}_{(k)}^*\|_* - (\|\mathbf{L}_{(k)}^*\|_* + \|P_k^\perp(\Delta)\|_* - \|P_k(\Delta)\|_*) \\ &= \|P_k(\Delta)\|_* - \|P_k^\perp(\Delta)\|_*. \end{aligned}$$

Then, we obtain

$$\begin{aligned} \|\mathcal{L}^*\|_{\text{t2nn}} - \|\hat{\mathcal{L}}\|_{\text{t2nn}} &\leq \left( \gamma \|P(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k(\Delta)\|_* \right) \\ &\quad - \left( \gamma \|P^\perp(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k^\perp(\Delta)\|_* \right). \end{aligned} \quad (22)$$

Using the definition of T2NN and triangular inequality yields

$$\begin{aligned} \|\Delta\|_{\text{t2nn}} &\leq \left( \gamma \|P(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k(\Delta)\|_* \right) \\ &\quad + \left( \gamma \|P^\perp(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k^\perp(\Delta)\|_* \right). \end{aligned} \quad (23)$$

Further using the setting  $\lambda \geq 2\sigma \|\mathfrak{X}^*(\mathbf{m}\xi)\|_{\text{t2nn}}^*$  yields Part (III),

$$\begin{aligned} \frac{1}{2} \|\mathfrak{X}(\Delta)\|_2^2 &\stackrel{(i)}{\leq} \frac{3\lambda}{2} \left( \gamma \|P(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k(\Delta)\|_* \right) \\ &\quad - \frac{\lambda}{2} \left( \gamma \|P^\perp(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k^\perp(\Delta)\|_* \right) \\ &\leq \frac{3\lambda}{2} \left( \gamma \|P(\Delta)\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k(\Delta)\|_* \right) \\ &\stackrel{(ii)}{\leq} \frac{3\lambda}{2} \left( \gamma \sqrt{2r_{\text{tb}}^*} \|P(\Delta)\|_{\text{F}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{2r_k^*} \|P_k(\Delta)\|_{\text{F}} \right) \\ &\stackrel{(iii)}{\leq} \frac{3\lambda}{2} \left( \gamma \sqrt{2r_{\text{tb}}^*} \|\Delta\|_{\text{F}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{2r_k^*} \|\Delta\|_{\text{F}} \right) \\ &= \frac{3\lambda}{2} \left( \gamma \sqrt{2r_{\text{tb}}^*} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{2r_k^*} \right) \|\Delta\|_{\text{F}}, \end{aligned}$$

where by combing (i) and  $\|\mathfrak{X}(\Delta)\|_2^2 \geq 0$ , Part (II) can be directly proved; inequality (ii) holds due to the compatible inequality of TNN and matrix nuclear norm, i.e.,  $\|\mathcal{T}\|_{\text{tnn}} \leq \sqrt{\text{rank}_{\text{tb}}(\mathcal{T})}\|\mathcal{T}\|_{\text{F}}$  [25] and  $\|\mathbf{T}\|_* \leq \sqrt{\text{rank}(\mathbf{T})}\|\mathbf{T}\|_{\text{F}}$  [27], and inequality (iii) holds because one can easily verify the facts that  $\|P(\Delta)\|_{\text{F}}^2 = \|\Delta\|_{\text{F}}^2 - \|P^\perp(\Delta)\|_{\text{F}}^2 \leq \|\Delta\|_{\text{F}}^2$  [25] and  $\|P_k(\Delta)\|_{\text{F}}^2 = \|\Delta_{(k)}\|_{\text{F}}^2 - \|P_k^\perp(\Delta)\|_{\text{F}}^2 \leq \|\Delta_{(k)}\|_{\text{F}}^2 = \|\Delta\|_{\text{F}}^2$  [27].

Note that inequality (20) gives an upper bound on the  $\|\mathfrak{X}(\Delta)\|_2$ , which can be seen as the ‘‘observed’’ error. However, we are more concerned about upper bounds on the error itself  $\|\Delta\|_{\text{F}}$  rather than its observed version. The following assumption builds a bridge between  $\|\mathfrak{X}(\Delta)\|_2$  and  $\|\Delta\|_{\text{F}}$ .

**Assumption 1 (RSC condition).** *The observation operator  $\mathfrak{X}(\cdot)$  is said to satisfy the Restricted Strong Convexity (RSC) condition with parameter  $\kappa$  if the following inequality holds:*

$$\|\mathfrak{X}(\mathcal{T})\|_2^2 \geq \kappa \|\mathcal{T}\|_{\text{F}}^2, \tag{24}$$

for any  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  belong to the restricted direction set,

$$\mathbf{C} := \left\{ \mathcal{T} \mid \gamma \|P^\perp(\mathcal{T})\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k^\perp(\mathcal{T})\|_* \leq 3 \left( \gamma \|P(\mathcal{T})\|_{\text{tnn}} + (1-\gamma) \sum_{k=1}^3 \alpha_k \|P_k(\mathcal{T})\|_* \right) \right\}. \tag{25}$$

Then, a straightforward combination of Proposition 1 and Assumption 1 leads to an deterministic bound on the estimation error.

**Theorem 1.** *By setting the regularization parameter  $\lambda \geq 2\sigma \|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$ , we have the following error bound for any solution  $\hat{\mathcal{L}}$  to Problem (15):*

$$\|\mathcal{L}^* - \hat{\mathcal{L}}\|_{\text{F}} \leq \frac{3\sqrt{2}}{\kappa} \lambda \left( \gamma \sqrt{r_{\text{tb}}^*} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{r_k^*} \right). \tag{26}$$

Note that we do not require information for distribution of the noise  $\xi$  in Theorem 1, which indicates that Theorem 1 provides a deterministic bound for general noise type. The bound on the right-hand side of Eq. 26 is in terms of the quantity,

$$\gamma \sqrt{r_{\text{tb}}^*} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{r_k^*},$$

which serves as a measure of structure complexity, reflecting the natural intuition that more complex structure causes larger error. The result is in consistent with the results for sum-of-norms-based estimators in [5, 22, 24, 28]. A more general analysis in [24, 28] indicates that the performance of sum-of-norms-based estimators are determined by all the structural complexities for a simultaneously structured signal, just as shown by the proposed bound (26).

## 4.2 Tensor Compressive Sensing

In this section, we consider tensor compressive sensing from random Gaussian design where  $\{\mathcal{X}_i\}$ ’s are random tensors with *i.*

*i.* standard Gaussian entries [22]. First, the RSC condition holds in random Gaussian design as shown in the following lemma. **Lemma 3 (RSC of random Gaussian design).** *If  $\mathfrak{X}(\cdot): \mathbb{R}^{d_1 \times d_2 \times d_3} \rightarrow \mathbb{R}^N$  is a random Gaussian design, then a version of the RSC condition is satisfied with probability at least  $1-2\exp(-N/32)$  as follows:*

$$\|\mathfrak{X}(\Delta)\| \geq \frac{\sqrt{N}}{4} \|\Delta\|_{\text{F}} - \frac{1}{16} \left( \frac{\sqrt{d_1 d_3} + \sqrt{d_2 d_3}}{\gamma} + \sum_{k=1}^3 \frac{\sqrt{d_k} + \sqrt{d_k}}{\alpha_k (1-\gamma)} \right) \|\Delta\|_{\text{t2nn}}, \tag{27}$$

for any tensor  $\Delta \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  in the restricted direction set  $\mathbf{C}$  whose definition is given in Eq. 25.

*Proof of Lemma 3.* The proof is analogous to that of Proposition 1 in [27]. The difference lies in how we lower bound the right hand side of (H.7) in [27], i.e.,

$$\mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \sup_{\mathbf{u} \in \mathbb{S}^{N-1}} Y_{\mathbf{u}, \Theta} \right], \tag{27a}$$

where  $\mathbb{S}^{N-1} := \{\mathbf{v} \in \mathbb{R}^N \mid \|\mathbf{v}\|_2 = 1\}$ ,  $\mathcal{R}(t) = \{\Theta \in \mathbb{R}^{d_1 \times d_2 \times d_3} \mid \|\Theta\|_{\text{F}} = 1, \|\Theta\|_{\text{t2nn}} \leq t\}$ , and

$$Y_{\mathbf{u}, \Theta} := \langle \mathbf{g}, \mathbf{u} \rangle + \langle \mathcal{G}, \Theta \rangle,$$

where random vector  $\mathbf{g} \in \mathbb{R}^N$  and random tensor  $\mathcal{G} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  are independent with *i. i.d.*  $\mathcal{N}(0, 1)$  entries.

We bound the quantity in Eq. 27 as follows:

$$\begin{aligned} \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \sup_{\mathbf{u} \in \mathbb{S}^{N-1}} Y_{\mathbf{u}, \Theta} \right] &= \mathbb{E} \left[ \sup_{\mathbf{u} \in \mathbb{S}^{N-1}} \langle \mathbf{g}, \mathbf{u} \rangle \right] + \mathbb{E} \left[ \inf_{\Theta \in \mathcal{R}(t)} \langle \mathcal{G}, \Theta \rangle \right] \\ &= \mathbb{E}[\|\mathbf{g}\|_2] - \mathbb{E} \left[ \sup_{\Theta \in \mathcal{R}(t)} \langle \mathcal{G}, \Theta \rangle \right] \\ &= \frac{1}{2} \sqrt{N} - t \mathbb{E}[\|\mathcal{G}\|_{\text{t2nn}}^*], \end{aligned} \tag{28}$$

where  $\|\mathcal{G}\|_{\text{t2nn}}^*$  can be bounded according to Lemma 4. The rest of the proof follows that of Proposition 1 in [27].

The remaining bound on  $\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$  is shown in the following Lemma.

**Lemma 4 (bound on  $\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$ ).** *Let  $\mathfrak{X}: \mathbb{R}^{d_1 \times d_2 \times d_3} \rightarrow \mathbb{R}^N$  be a random Gaussian design. With high probability, the quantity  $\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$  is concentrated around its mean, which can be bounded as follows:*

$$\mathbb{E}[\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*] \leq c\sqrt{N} \left( \frac{\sqrt{d_1 d_3} + \sqrt{d_2 d_3}}{\gamma} + \sum_{k=1}^3 \frac{\sqrt{d_k} + \sqrt{d_k}}{\alpha_k (1-\gamma)} \right). \tag{29}$$

*Proof.* Since  $\xi_m$ ’s are *i. i.d.*  $\mathcal{N}(0, 1)$  variables, we have

$$\|\xi\|_2 \leq 2\sqrt{N} \tag{29a}$$

with high probability according to Proposition 8.1 in [29].

For  $k = 1, 2, 3$ , let  $\mathbf{X}^*(\xi)_{(k)}$  be the mode-*k* unfolding of random tensor  $\mathfrak{X}^*(\xi)$ . A direct use of Lemma C.1 in [27] leads to

$$\mathbb{E}[\|\mathbf{X}^*(\xi)_{(k)}\|] \leq c_0 \sqrt{N} (\sqrt{d_k} + \sqrt{d_k}) \tag{30}$$

with high probability. A similar argument of Lemma C.1 in [27] also yields

$$\mathbb{E}[\|\hat{\mathbf{x}}^*(\xi)\|] \leq c_1 \sqrt{N} (\sqrt{d_1 d_3} + \sqrt{d_2 d_3}) \quad (31)$$

with high probability. Combining Eqs 30, 31, we can complete the proof.

Then, the non-asymptotic error bound is obtained finally as follows.

**Theorem 2 (non-asymptotic error bound).** *Under the random Gaussian design setup, there are universal constants  $c_3, c_4$ , and  $c_5$  such that for a sample size  $N$  greater than*

$$c_3 \left( \frac{\sqrt{d_1 d_3} + \sqrt{d_2 d_3}}{\gamma} + \sum_{k=1}^3 \frac{\sqrt{d_k} + \sqrt{d_{3k}}}{\alpha_k (1-\gamma)} \right)^2 \times \left( \gamma \sqrt{r_{\text{tb}}^*} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{r_k^*} \right)^2$$

and any solution to Problem (15) with regularization parameter

$$\lambda = c_4 \sigma \sqrt{N} \left( \frac{\sqrt{d_1 d_3} + \sqrt{d_2 d_3}}{\gamma} + \sum_{k=1}^3 \frac{\sqrt{d_k} + \sqrt{d_{3k}}}{\alpha_k (1-\gamma)} \right),$$

then we have

$$\|\Delta\|_{\text{F}}^2 \leq c_5 \sigma^2 \left( \frac{\sqrt{d_1 d_3} + \sqrt{d_2 d_3}}{\gamma} + \sum_{k=1}^3 \frac{\sqrt{d_k} + \sqrt{d_{3k}}}{\alpha_k (1-\gamma)} \right)^2 \frac{\left( \gamma \sqrt{r_{\text{tb}}^*} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{r_k^*} \right)^2}{N}, \quad (32)$$

which holds with high probability.

To understand the proposed bound, we consider the three-way cubical tensor  $\mathcal{L}^* \in \mathbb{R}^{d \times d \times d}$  with regularization weights  $\gamma = (1-\gamma)\alpha_1 = (1-\gamma)\alpha_2 = (1-\gamma)\alpha_3 = 1/4$ . Then, the bound in Eq. 52 is simplified to the following element-wise error:

$$\frac{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2}{d^3} \leq O \left( \sigma^2 \cdot \frac{1}{N} \cdot \left( \sqrt{\frac{r_{\text{tb}}^*}{d}} + \sum_{k=1}^3 \sqrt{\frac{r_k^*}{d}} \right)^2 \right), \quad (33)$$

which means the estimation error is controlled by the tubal rank and Tucker rank of  $\mathcal{L}^*$  simultaneously. From the right-hand side of Eq. 33, it can be seen that the more observations (i.e., the larger  $N$ ), the smaller the error; it is also reflected that larger tensors with more complex structures will lead to larger errors. The interpretation is consistent with our intuition.

**Equation 33** also indicates the sample size  $N$  should satisfy

$$N \geq \Omega \left( \left( \sqrt{r_{\text{tb}}^*} + \sum_{k=1}^3 \sqrt{r_k^*} \right)^2 d^2 \right) \quad (34)$$

for approximate tensor sensing.

Another interesting result is that by setting the noise level  $\sigma = 0$  in Eq. 33, the upper bound reaches 0, which means the proposed estimator can exactly recover the unknown truth  $\mathcal{L}^*$  in the noiseless setting.

### 4.3 Noisy Tensor Completion

For noisy tensor completion, we consider a slightly modified estimator,

$$\hat{\mathcal{L}} \in \underset{\|\mathcal{L}\|_{\infty} \leq \mathbf{a}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}(\mathcal{L})\|_2^2 + \lambda \|\mathcal{L}\|_{\text{t2nn}}, \quad (35)$$

where  $\mathbf{ma} > 0$  is a known constant constraining the magnitude of entries in  $\mathcal{L}^*$ . The constraint  $\|\mathcal{L}\|_{\infty} \leq \mathbf{a}$  is very mild because real signals are all of limited magnitude, e.g., the intensity of pixels in visual light images cannot be greater than 255. The constraint also provides with theoretical continence in excluding the “spiky” tensors while controlling the identifiability of  $\mathcal{L}^*$ . Similar “non-spiky” constraints are also considered in related work [6, 16, 30].

We consider noisy tensor completion under uniform sampling in this section.

**Assumption 2 (uniform sampling scheme).** *The design tensors  $\{\mathcal{X}_i\}$  are i.i.d. random tensor bases drawn from uniform distribution  $\Pi$  on the set,*

$$\left\{ \mathbf{e}_i \circ \mathbf{e}_j \circ \mathbf{e}_k : \forall (i, j, k) \in [d_1] \times [d_2] \times [d_3] \right\}.$$

Recall that Proposition 1 in Section 4.1 gives an upper bound on the “observed part” of the estimation error  $\|\mathbf{x}(\Delta)\|_{\text{F}}$ . As our goal is to establish a bound on  $\|\Delta\|_{\text{F}}$ , we then connect  $\|\mathbf{x}(\Delta)\|_{\text{F}}$  with  $\|\Delta\|_{\text{F}}$  by quantifying the probability of the following RSC property of the sampling operator  $\mathbf{x}$ :

$$\frac{1}{N} \|\mathbf{x}(\Delta)\|_{\text{F}}^2 \geq \frac{1}{2d_1 d_2 d_3} \|\Delta\|_{\text{F}}^2 - \text{an intercept term},$$

when the error tensor  $\Delta$  belongs to some set  $\mathbf{mC}(\beta, \mathbf{mr})$  defined as

$$\mathbf{C}(\beta, r) := \left\{ \Delta \in \mathbb{R}^{d_1 \times d_2 \times d_3} \mid \|\Delta\|_{\infty} \leq 1, \frac{\|\Delta\|_{\text{F}}^2}{d_1 d_2 d_3} \geq \beta, \|\Delta\|_{\text{t2nn}} \leq \left( \gamma \sqrt{r} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{r_k} \right) \|\Delta\|_{\text{F}} \right\}, \quad (36)$$

where  $\beta = \sqrt{\frac{64 \log \bar{d}}{N \log(6/5)}}$  is an F-norm tolerance parameter and  $\mathbf{mr} = (r, r_1, r_2, r_3)$  is a rank parameter whose values will be specified in the sequel.

**Lemma 5 (RSC condition under uniform sampling).** *For any  $\Delta \in \mathbf{C}(\beta, r)$ , it holds with probability at least  $1 - (d_1 d_3 + d_2 d_3)^{-1}$  that*

$$\frac{1}{N} \|\mathbf{x}(\Delta)\|_{\text{F}}^2 \geq \frac{\|\Delta\|_{\text{F}}^2}{2d_1 d_2 d_3} - \frac{44d_1 d_2 d_3}{N^2} \mathbb{E}[\|\mathbf{x}^*(\epsilon)\|_{\text{t2nn}}^*]^2 \times \left( \gamma \sqrt{r} + (1-\gamma) \sum_{k=1}^3 \alpha_k \sqrt{r_k} \right)^2, \quad (37)$$

where  $e$  is the base of the natural logarithm, and the entries  $\epsilon_i$  of vector  $\epsilon \in \mathbb{R}^N$  are i. i.d. Rademacher random variables.

Before proving Lemma 5, we first define a subset of  $\mathbf{C}(\beta, r)$  by upper bounding the F-norm of any element  $\Delta$  in it,



$$\mathbf{B}(r, T) := \left\{ \Delta: \Delta \in \mathbf{C}(\beta, r), \frac{\|\Delta\|_{\mathbb{F}}^2}{d_1 d_2 d_3} \leq T \right\},$$

and a quantity

$$Z_T := \sup_{\Delta \in \mathbf{B}(r, T)} \left| \frac{\|\mathfrak{X}(\Delta)\|_2^2}{N} - \frac{\|\Delta\|_{\mathbb{F}}^2}{d_1 d_2 d_3} \right|,$$

which is the maximal absolute deviation of  $N^{-1}\|\mathfrak{X}(\Delta)\|_2^2$  from its expectation  $(d_1 d_2 d_3)^{-1}\|\Delta\|_{\mathbb{F}}^2$  in  $\mathbf{B}(r, T)$ . Lemma 6 shows the concentration behavior of  $Z_T$ .

Lemma 6 (concentration of  $Z_T$ ). *There exists a constant  $c_0$  such that*

$$\mathbb{P} \left[ Z_T \geq \frac{5}{12} T - \frac{44 d_1 d_2 d_3}{N^2} \mathbb{E} [\|\mathfrak{X}^*(\epsilon)\|]^2 (\gamma \sqrt{r} + (1 - \gamma) \times \sum_k \alpha_k \sqrt{r_k})^2 \right] \leq \exp(-c_0 N T^2).$$

*Proof of Lemma 6.* The proof is similar to that of Lemma 10 in [31]. The difference lies in the step of symmetrization arguments. Note that for any  $\Delta \in \mathbf{B}(r, T)$ , it holds that

$$\|\Delta\|_{\text{tnn}} \leq \sqrt{d_1 d_2 d_3} T (\gamma \sqrt{r} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k}),$$

which indicates

$$\begin{aligned} \mathbb{E}[Z_T] &\leq 8 \mathbb{E} \left[ \sup_{\Delta \in \mathbf{B}(r, T)} \langle \mathfrak{X}^*(\epsilon), \Delta \rangle \right] \\ &\leq 8 \mathbb{E} \left[ \sup_{\Delta \in \mathbf{B}(r, T)} \frac{1}{d_3} \|\mathfrak{X}^*(\epsilon)\| \|\Delta\|_{\text{tnn}} \right] \\ &\leq 8 \sqrt{d_1 d_2 d_3} T \mathbb{E} [\|\mathfrak{X}^*(\epsilon)\|] (\gamma \sqrt{r} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k}). \end{aligned}$$

Then, Lemma 5 can be proved by using the peeling argument [30].

*Proof of Lemma 5.* For any positive integer  $l$ , we define disjoint subsets of  $\mathbf{C}(\beta, r)$  as

$$\mathbf{D}(r, l) := \left\{ \Delta: \Delta \in \mathbf{C}(\beta, r), \beta \rho^{l-1} \leq \frac{\|\Delta\|_{\mathbb{F}}^2}{d_1 d_2 d_3} \leq \beta \rho^l \right\}$$

with constants  $\rho = \frac{6}{5}$  and  $\beta = \sqrt{\frac{64 \log \tilde{d}}{N \log \rho}}$ . Let  $D = d_1 d_2 d_3$  for simplicity, and define the event

$$\begin{aligned} \mathfrak{G} := & \left\{ \exists \Delta \in \mathbf{C}(\beta, r), \text{ s.t. } \left| \frac{\|\mathfrak{X}(\Delta)\|_2^2}{N} - \frac{\|\Delta\|_{\mathbb{F}}^2}{D} \right| \geq \frac{\|\Delta\|_{\mathbb{F}}^2}{2D} \right. \\ & \left. + \frac{44D}{N^2} \mathbb{E}^2 [\|\mathfrak{X}^*(\epsilon)\|] \left( \gamma \sqrt{r} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k} \right)^2 \right\} \end{aligned}$$

and its sub-events for any  $l \in \mathbb{N}_+$ ,

$$\begin{aligned} \mathfrak{G}_l := & \left\{ \exists \Delta \in \mathbf{D}(r, l), \text{ s.t. } \left| \frac{\|\mathfrak{X}(\Delta)\|_2^2}{N} - \frac{\|\Delta\|_{\mathbb{F}}^2}{D} \right| \geq \frac{5}{12} \beta \rho^l \right. \\ & \left. + \frac{44D}{N^2} \mathbb{E}^2 [\|\mathfrak{X}^*(\epsilon)\|] \left( \gamma \sqrt{r} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k} \right)^2 \right\}. \end{aligned}$$

Note that Lemma 6 implies that

$$\begin{aligned} \mathbb{P}[\mathfrak{G}_l] &= \mathbb{P}_{\Delta \in \mathbf{C}(r, l)} \left[ Z_T \geq \frac{5}{12} \beta \rho^l + \frac{44D}{N^2} \mathbb{E}^2 [\|\mathfrak{X}^*(\epsilon)\|] \right. \\ & \quad \left. \times \left( \gamma \sqrt{r} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k} \right)^2 \right] \\ &\leq \exp(-c_0 N \beta^2 \rho^{2l}). \end{aligned}$$

Thus, we have

$$\begin{aligned} \mathbb{P}[\mathfrak{G}] &\leq \sum_{l=1}^{\infty} \mathbb{P}[\mathfrak{G}_l] \leq \sum_{l=1}^{\infty} \exp(-c_0 N \beta^2 \rho^{2l}) \\ &\leq \sum_{l=1}^{\infty} \exp(-2c_0 N \beta^2 l \log \rho) \\ &\leq \frac{\exp(-c_0 \beta^2 \log \rho)}{1 - \exp(-c_0 \beta^2 \log \rho)}. \end{aligned}$$

Recall that  $\beta = \sqrt{\frac{64 \log \tilde{d}}{\log(6/5)N}}$ , then  $\mathbb{P}[\mathfrak{G}] \leq 2/\tilde{d}$ , which leads to the result of Lemma 5.

Based on the RSC condition in Lemma 5, we are able to give an upper bound on the estimation error  $\|\Delta\|_{\mathbb{F}}$  in the following proposition. Proposition 2. *With parameter  $\lambda \geq 2\sigma \|\mathfrak{X}^*(\xi)\|_{\text{tnn}}^*$ , the estimation error satisfies*

$$\begin{aligned} \frac{\|\Delta\|_{\mathbb{F}}^2}{d_1 d_2 d_3} &\leq \max \left\{ c_1 \frac{d_1 d_3 d_3}{N^2} (\lambda^2 + \mathbf{a}^2 \mathbb{E} [\|\mathfrak{X}^*(\epsilon)\|]^2) \right. \\ & \quad \left. \times \left( \gamma \sqrt{r_{\text{tb}}} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k^*} \right)^2, c_2 \mathbf{a}^2 \sqrt{\frac{\log \tilde{d}}{N}} \right\} \end{aligned} \quad (38)$$

with probability at least  $1 - 2(d_1 d_3 + d_2 d_3)^{-1}$ .

*Proof of Proposition 2.* A direct consequence of property (II) in Proposition 1 and the triangular inequality is that the error tensor  $\Delta$  satisfies

$$\|\Delta\|_{\text{tnn}} \leq \left( \gamma \sqrt{32r_{\text{tb}}^*} + (1 - \gamma) \sum_{k=1}^3 \alpha_k \sqrt{32r_k^*} \right) \|\Delta\|_{\mathbb{F}}. \quad (39)$$

Since  $\|\hat{\mathcal{L}}\|_{\infty} < \mathbf{a}$  and  $\|\mathcal{L}^*\|_{\infty} < \mathbf{a}$ , we also have  $\|\Delta\|_{\infty} \leq \|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\infty} \leq \|\hat{\mathcal{L}}\|_{\infty} + \|\mathcal{L}^*\|_{\infty} < 2\mathbf{a}$ .

Let  $\mathbf{r}^* = (r_{\text{tb}}^*, r_1^*, r_2^*, r_3^*)^T$  denote the rank complexity of the underlying tensor  $\mathcal{L}^*$ . By discussing whether tensor  $\frac{\Delta}{2\mathbf{a}}$  is in set  $\mathbf{C}(\beta, 32\mathbf{r}^*)$ , we consider the following cases.

*Case 1:* If  $\frac{\Delta}{2\mathbf{a}} \notin \mathbf{C}(\beta, 32\mathbf{r}^*)$ , then from the definition of set  $\mathbf{C}(\beta, r)$ , we have

$$\frac{\|\Delta\|_{\mathbb{F}}^2}{d_1 d_2 d_3} \leq 4\mathbf{a}^2 \sqrt{\frac{64 \log \tilde{d}}{\log(6/5)N}}. \quad (40)$$

*Case 2:* If  $\frac{\Delta}{2\mathbf{a}} \in \mathbf{C}(\beta, 32\mathbf{r}^*)$ , then by Proposition 1 and Lemma 5, we have

$$\begin{aligned} \frac{\|\Delta\|_{\mathbb{F}}^2}{2\mathbf{a}} &\leq \frac{44 \cdot 32D}{N^2} \mathbb{E}^2 [\|\mathfrak{X}^*(\epsilon)\|] \times \left( \gamma \sqrt{r_{\text{tb}}} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k^*} \right)^2 \\ &\leq \frac{3\sqrt{2}\lambda}{N} \left( \gamma \sqrt{r_{\text{tb}}} + (1 - \gamma) \sum_k \alpha_k \sqrt{r_k^*} \right) \left\| \frac{\Delta}{2\mathbf{a}} \right\|_{\mathbb{F}} \end{aligned} \quad (41)$$

with probability at least  $1 - 2(d_1 d_3 + d_2 d_3)^{-1}$ .

By performing some algebra (like the proof of Theorem 3 in [30]), we have

$$\frac{\|\Delta\|_F^2}{d_1 d_2 d_3} \leq C \frac{d_1 d_2 d_3}{N^2} (\lambda^2 + \mathfrak{a}^2 \mathbb{E}^2[\|\mathfrak{X}^*(\epsilon)\|]) \times \left( \gamma \sqrt{r_{\text{tb}}^*} + (1-\gamma) \sum_k \alpha_k \sqrt{r_k^*} \right)^2 \quad (42)$$

Combining Case 1 and Case 2, we obtain the result of Proposition 2.

According to Proposition 2 and Lemma 5, it remains to bound  $\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$  and  $\mathbb{E}[\|\mathfrak{X}^*(\epsilon)\|_{\text{t2nn}}^*]$ . The following lemmas give their bounds respectively. As the noise variables  $\{\xi_i\}$  are *i. i. d.* standard Gaussian, it belongs to the sub-exponential distribution [32], and thus, there exists a constant  $\varrho$  as the smallest number satisfying [30]

$$\max_{i \leq N} \mathbb{E} \left[ e^{\frac{|\xi_i|}{\varrho}} \right] \leq e. \quad (43)$$

Suppose the sample complexity  $N$  in noisy tensor completion satisfies

$$\begin{aligned} N &\geq \max\{d_1 d_3 \vee d_2 d_3, \max_k (d_k \vee d_{\vee k})\} \\ N &\geq 2(d_1 \wedge d_2) \varrho^2 \log^2(\varrho \sqrt{d_1 \wedge d_2}) \log(d_1 d_3 + d_2 d_3) \\ N &\geq \max_{k \leq 3} 2(d_k \wedge d_{\vee k}) \varrho^2 \log^2(\varrho \sqrt{d_k \wedge d_{\vee k}}) \log(d_k + d_{\vee k}) \\ N &\geq 2(d_1 \wedge d_2) \log^2(\sqrt{d_1 \wedge d_2}) \log(d_1 d_3 + d_2 d_3) \\ N &\geq \max_{k \leq 3} 2(d_k \wedge d_{\vee k}) \log^2(\sqrt{d_k \wedge d_{\vee k}}) \log(d_k + d_{\vee k}). \end{aligned} \quad (44)$$

Then, we have the following Lemma 7 and Lemma 8 to bound  $\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^*$  and  $\mathbb{E}[\|\mathfrak{X}^*(\epsilon)\|_{\text{t2nn}}^*]$ .

Lemma 7. *Under the sample complexity of noisy tensor completion in Eq. 44, it holds with probability at least  $1 - (d_1 d_3 + d_2 d_3)^{-1} - \sum_k (d_k + d_{\vee k})^{-1}$  that*

$$\|\mathfrak{X}^*(\xi)\|_{\text{t2nn}}^* \leq C_\varrho \sqrt{N} \left( \frac{1}{\gamma} \sqrt{\frac{\log(d_1 d_3 + d_2 d_3)}{d_1 \wedge d_2}} + \frac{1}{(1-\gamma)} \times \sum_{k=1}^3 \frac{1}{\alpha_k} \sqrt{\frac{\log(d_k + d_{\vee k})}{d_k \wedge d_{\vee k}}} \right), \quad (45)$$

where  $C_\varrho$  is a constant dependent on the  $\varrho$  that is defined in Eq. 43.

Proof of Lemma 7. The proof can be straightforwardly obtained by adopting the upper bound of the dual T2NN norm in Lemma 2 and Lemma 5 in the supplementary material of [25], and Lemma 5 in [30] as follows:

- First, Lemma 5 in the supplementary material of [25] shows that letting  $N \geq d_1 d_3 \vee d_2 d_3$  and  $N \geq 2(d_1 \wedge d_2) \varrho^2 \log^2(\varrho \sqrt{d_1 \wedge d_2}) \log(d_1 d_3 + d_2 d_3)$ , then it holds with probability at least  $1 - (d_1 d_3 + d_2 d_3)^{-1}$  that

$$\|\mathfrak{X}^*(\xi)\| \leq C_\varrho \sqrt{N(d_1 \wedge d_2)^{-1} \log(d_1 d_3 + d_2 d_3)}. \quad (46)$$

- For  $k = 1, 2, 3$ , let  $\mathbf{X}^*(\xi)_{(k)}$  be the mode- $k$  unfolding of random tensor  $\mathfrak{X}^*(\xi)$ . Then, Lemma 5 in [30]

indicates that letting  $N \geq d_k \vee (d_{\vee k})$  and  $N \geq 2(d_k \wedge (d_{\vee k})) \varrho^2 \log^2(\varrho \sqrt{d_k \wedge d_{\vee k}}) \log(d_k + d_{\vee k})$ , then it holds with probability at least  $1 - (d_k + d_{\vee k})^{-1}$  that

$$\|\mathbf{X}^*(\xi)_{(k)}\| \leq C'_\varrho \sqrt{N(d_k \wedge d_{\vee k})^{-1} \log(d_k + d_{\vee k})}. \quad (47)$$

Then, combining Eq. 46 and 47 and using union bound, Eq. 45 can be obtained.

Lemma 8. *Under the sample complexity of noisy tensor completion in Eq. 44, it holds that*

$$\mathbb{E}[\|\mathfrak{X}^*(\epsilon)\|_{\text{t2nn}}^*] \leq C \sqrt{N} \left( \frac{1}{\gamma} \sqrt{\frac{\log(d_1 d_3 + d_2 d_3)}{d_1 \wedge d_2}} + \frac{1}{(1-\gamma)} \times \sum_{k=1}^3 \frac{1}{\alpha_k} \sqrt{\frac{\log(d_k + d_{\vee k})}{d_k \wedge d_{\vee k}}} \right). \quad (48)$$

Proof of Lemma 8. Similar to the proof of Lemma 7, the proof can be straightforwardly obtained by adopting the upper bound of the dual T2NN norm in Lemma 2 and Lemma 6 in the supplementary material of [25], and Lemma 6 in [30].

- First, Lemma 6 in the supplementary material of [25] shows that letting  $N \geq d_1 d_3 \vee d_2 d_3$  and  $N \geq 2(d_1 \wedge d_2) \log^2(d_1 \wedge d_2) \log(d_1 d_3 + d_2 d_3)$ , then, the following inequality holds:

$$\mathbb{E}[\|\mathfrak{X}^*(\epsilon)\|] \leq C_2 \sqrt{N(d_1 \wedge d_2)^{-1} \log(d_1 d_3 + d_2 d_3)}. \quad (49)$$

- For  $k = 1, 2, 3$ , let  $\mathbf{X}^*(\epsilon)_{(k)}$  be the mode- $k$  unfolding of random tensor  $\mathfrak{X}^*(\epsilon)$ . Then, Lemma 6 in [30] indicates that letting  $N \geq d_k \vee d_{\vee k}$  and  $N \geq 2(d_k \wedge d_{\vee k}) \log^2(d_k \wedge d_{\vee k}) \log(d_k + d_{\vee k})$ , then, the following inequality holds:

$$\mathbb{E}[\|\mathbf{X}^*(\epsilon)_{(k)}\|] \leq C'_2 \sqrt{N(d_k \wedge d_{\vee k})^{-1} \log(d_k + d_{\vee k})}. \quad (50)$$

Then, Eq. 48 can be obtained by combining Eqs. 49 and 50.

Further combining Lemma 7, Lemma 8, and Proposition 2, we arrive at an upper bound on the estimation error in the follow theorem.

Theorem 3. *Suppose Assumption 2 is satisfied and  $\|\mathcal{L}^*\|_\infty \leq \mathfrak{a}$ . Let the sample size  $N$  satisfies Eq. 44. By setting*

$$\lambda = C_\varrho \sigma \sqrt{N} \left( \frac{1}{\gamma} \sqrt{\frac{\log(d_1 d_3 + d_2 d_3)}{d_1 \wedge d_2}} + \frac{1}{(1-\gamma)} \sum_{k=1}^3 \frac{1}{\alpha_k} \sqrt{\frac{\log(d_k + d_{\vee k})}{d_k \wedge d_{\vee k}}} \right), \quad (51)$$

the estimation error of any estimator  $\hat{\mathcal{L}}$  defined in Problem (35) can be upper bounded as follows:

$$\begin{aligned} \frac{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2}{d_1 d_2 d_3} &\leq c_2 \max \left\{ a^2 \sqrt{\frac{\log \tilde{d}}{N}}, \frac{d_1 d_2 d_3 (\sigma^2 \vee \mathbf{a}^2)}{N} \right. \\ &\quad \times \left( \gamma \sqrt{r^*} + (1-\gamma) \sum_k \alpha_k \sqrt{r_k^*} \right)^2 \\ &\quad \cdot \left( \frac{1}{\gamma} \sqrt{\frac{\log(d_1 d_3 + d_2 d_3)}{d_1 \wedge d_2}} + \frac{1}{(1-\gamma)} \right. \\ &\quad \left. \left. \times \sum_{k=1}^3 \frac{1}{\alpha_k} \sqrt{\frac{\log(d_k + d_{\vee k})}{d_k \wedge d_{\vee k}}} \right)^2 \right\}. \end{aligned} \quad (52)$$

with probability at least  $1 - 3(d_1 d_3 + d_2 d_3)^{-1} - \sum_k (d_k + d_{\vee k})^{-1}$ .

To understand the proposed bound in Theorem 3, we consider the three-way cubical tensor  $\mathcal{L}^* \in \mathbb{R}^{d \times d \times d}$  with regularization weights  $\gamma = (1-\gamma)\alpha_1 = (1-\gamma)\alpha_2 = (1-\gamma)\alpha_3 = 1/4$ . Then, the bound in Eq. 52 is simplified to the following element-wise error:

$$\frac{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2}{d^3} \leq O\left(\frac{d^3}{N} \cdot (\sigma \vee \mathbf{a})^2 \cdot \left(\sqrt{\frac{r^*}{d}} + \sum_{k=1}^3 \sqrt{\frac{r_k^*}{d}}\right)^2 \log d\right), \quad (53)$$

which means the estimation error is controlled by the tubal rank and Tucker rank of  $\mathcal{L}^*$  simultaneously. Equation 53 also indicates that the sample size  $N$  should satisfy

$$N \geq \Omega\left(\left(\sqrt{r^*} + \sum_{k=1}^3 \sqrt{r_k^*}\right)^2 d^2 \log d\right) \quad (54)$$

for approximate tensor completion.

## 5 OPTIMIZATION ALGORITHM

The ADMM framework [33] is applied to solve the proposed model. Adding auxiliary variables  $\mathcal{K}$  and  $\mathcal{T}^1, \mathcal{T}^2, \mathcal{T}^3$  to Problem (15) yields an equivalent formulation,

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{K}, \{\mathcal{T}^k\}_k} & \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2^2 + \lambda \gamma \|\mathcal{K}\|_{\text{tnn}} + \lambda(1-\gamma) \\ & \times \sum_{k=1}^3 \alpha_k \|\mathbf{T}_{(k)}^k\|_* \\ \text{s.t.} & \quad \mathcal{K} = \mathcal{L}; \quad \mathcal{T}^k = \mathcal{L}, \quad k = 1, 2, 3. \end{aligned} \quad (55)$$

To solve Problem (55), an ADMM-based algorithm is proposed. First, the augmented Lagrangian is

$$\begin{aligned} L_p(\mathcal{L}, \mathcal{K}, \{\mathcal{T}^k\}_k, \mathcal{A}, \{\mathcal{B}^k\}_k) &= \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2^2 + \lambda \gamma \|\mathcal{K}\|_{\text{tnn}} \\ &+ \langle \mathcal{A}, \mathcal{K} - \mathcal{L} \rangle + \frac{\rho}{2} \|\mathcal{K} - \mathcal{L}\|_{\text{F}}^2 \\ &+ \lambda(1-\gamma) \sum_{k=1}^3 \left( \alpha_k \|\mathbf{T}_{(k)}^k\|_* + \langle \mathcal{B}^k, \mathcal{T}^k - \mathcal{L} \rangle + \frac{\rho}{2} \|\mathcal{T}^k - \mathcal{L}\|_{\text{F}}^2 \right), \end{aligned} \quad (56)$$

where tensors  $\mathcal{A}$  and  $\{\mathcal{B}^k\}_k$  are the dual variables.

The primal variables  $\mathcal{L}, \mathcal{K}$ , and  $\mathcal{T}^k$  can be divided into two blocks: The first block has one tensor variable  $\mathcal{L}$ , whereas the second block consists of four variables  $\mathcal{K}$  and  $\mathcal{T}^k$ 's. We use the minimization scheme of ADMM to update the two blocks alternatively after the  $t$ th iteration ( $t = 0, 1, \dots$ ):

*Update the first block  $\mathcal{L}$ :* We update  $\mathcal{L}$  by solving following  $\mathcal{L}$ -subproblem with all the other variables fixed:

$$\begin{aligned} \mathcal{L}^{t+1} &= \underset{\mathcal{L}}{\text{argmin}} L_p(\mathcal{L}, \mathcal{K}^t, \{(\mathcal{T}^k)^t\}_k, \mathcal{A}^t, \{(\mathcal{B}^k)^t\}_k) \\ &= \underset{\mathcal{L}}{\text{argmin}} \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2^2 + \frac{\rho}{2} \|\mathcal{K}^t + \rho^{-1} \mathcal{A}^t - \mathcal{L}\|_{\text{F}}^2 \\ &\quad + \sum_{k=1}^3 \frac{\rho}{2} \|(\mathcal{T}^k)^t + \rho^{-1} (\mathcal{B}^k)^t - \mathcal{L}\|_{\text{F}}^2. \end{aligned}$$

By taking derivative with respect to  $\mathcal{L}$  and setting the derivative to zero, we obtain the following equation:

$$\begin{aligned} \mathfrak{X}^*(\mathfrak{X}(\mathcal{L}) - \mathbf{y}) + \rho(\mathcal{L} - \mathcal{K}^t - \rho^{-1} \mathcal{A}^t) \\ + \sum_{k=1}^3 \rho(\mathcal{L} - (\mathcal{T}^k)^t - \rho^{-1} (\mathcal{B}^k)^t) = \mathbf{0}. \end{aligned}$$

Solving the above equation yields

$$\begin{aligned} \mathcal{L}^{t+1} &= (\mathfrak{X}^* \mathfrak{X} + 4\rho \mathbb{I})^{-1} \left( \mathfrak{X}^*(\mathbf{y}) + \rho \mathcal{K}^t + \mathcal{A}^t \right. \\ &\quad \left. + \sum_{k=1}^3 (\rho (\mathcal{T}^k)^t + (\mathcal{B}^k)^t) \right), \end{aligned} \quad (57)$$

where  $\mathbb{I}(\cdot)$  is the identity operator.

*Update the second block  $(\mathcal{K}, \{\mathcal{T}^k\})$ :* We update  $\mathcal{K}$  and  $\{\mathcal{T}^k\}$  in parallel by keeping all the other variables fixed. First,  $\mathcal{K}$  is updated by solving the  $\mathcal{K}$ -subproblem,

$$\begin{aligned} \mathcal{K}^{t+1} &= \underset{\mathcal{K}}{\text{argmin}} L_p(\mathcal{L}^{t+1}, \mathcal{K}, \{(\mathcal{T}^k)^t\}_k, \mathcal{A}^t, \{(\mathcal{B}^k)^t\}_k) \\ &= \underset{\mathcal{K}}{\text{argmin}} \lambda \gamma \|\mathcal{K}\|_{\text{tnn}} + \langle \mathcal{A}^t, \mathcal{K} - \mathcal{L}^{t+1} \rangle + \frac{\rho}{2} \|\mathcal{K} - \mathcal{L}^{t+1}\|_{\text{F}}^2 \\ &= \text{Prox}_{\rho^{-1} \lambda \gamma}^{\|\cdot\|_{\text{tnn}}}(\mathcal{L}^{t+1} - \rho^{-1} \mathcal{A}^t), \end{aligned} \quad (58)$$

where  $\text{Prox}_\tau^{\|\cdot\|_{\text{tnn}}}(\cdot)$  is the proximal operator of TNN given in Lemma 9.

Then,  $\mathcal{T}^k$  is updated by solving the  $\mathcal{T}^k$ -subproblem ( $k = 1, 2, 3$ ),

$$\begin{aligned} (\mathcal{T}^k)^{t+1} &= \underset{\mathcal{T}^k}{\text{argmin}} L_p(\mathcal{L}^{t+1}, \mathcal{K}^t, \{(\mathcal{T}^k)^t\}_k, \mathcal{A}^t, \{(\mathcal{B}^k)^t\}_k) \\ &= \underset{\mathcal{T}^k}{\text{argmin}} \lambda \alpha_k (1-\gamma) \|\mathbf{T}_{(k)}^k\|_* + \langle (\mathcal{B}^k)^t, \mathcal{T}^k - \mathcal{L}^{t+1} \rangle + \frac{\rho}{2} \|\mathcal{T}^k - \mathcal{L}^{t+1}\|_{\text{F}}^2 \\ &= \mathfrak{F}_k \left( \text{Prox}_{\rho^{-1} \lambda \alpha_k (1-\gamma)}^{\|\cdot\|_*} (\mathcal{L}^{t+1} - \rho^{-1} (\mathcal{B}^k)^t) \right), \end{aligned} \quad (59)$$

where  $\mathfrak{F}_k(\cdot): \mathbb{R}^{d_k \times d_{\vee k}} \rightarrow \mathbb{R}^{d_1 \times d_2 \times d_3}$  is the folding function to reshape a mode- $k$  matricization to its original tensor format and  $\text{Prox}_\tau^{\|\cdot\|_*}(\cdot)$  is the proximal operator of matrix nuclear norm given in Lemma 10.

Lemma 9 (proximal operator of TNN [34]). *Let tensor  $\mathcal{T}_0 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  with  $t$ -SVD  $\mathcal{T}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ , where  $\mathcal{U} \in \mathbb{R}^{d_1 \times r \times d_3}$  and  $\mathcal{V} \in \mathbb{R}^{d_2 \times r \times d_3}$  are orthogonal tensors and  $\mathcal{S} \in \mathbb{R}^{r \times r \times d_3}$  is the  $f$ -diagonal tensor of singular tubes. Then, the*

proximal operator of function  $\|\cdot\|_{\text{ttn}}$  at point  $\mathcal{T}_0$  with parameter  $\tau$  can be computed as follows:

$$\begin{aligned} \text{Prox}_{\tau}^{\|\cdot\|_{\text{ttn}}}(\mathcal{T}_0) &:= \underset{\mathcal{T}}{\text{argmin}} \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_{\text{F}}^2 + \tau\|\mathcal{T}\|_{\text{ttn}} \\ &= \mathbf{U} * \text{ifft}_3(\max(\text{fft}_3(\mathcal{S}) - \tau, 0)) * \mathbf{V}^{\text{T}}, \end{aligned}$$

where  $\text{fft}_3(\cdot)$  and  $\text{ifft}_3(\cdot)$  denote the operations of fast DFT and fast inverse DFT on all the tubes of a given tensor, respectively.

Lemma 10 (proximal operator of the matrix nuclear norm [35]). Let tensor  $\mathbf{T}_0 \in \mathbb{R}^{d_1 \times d_2}$  with SVD  $\mathbf{T}_0 = \mathbf{U}\mathbf{S}\mathbf{V}^{\text{T}}$ , where  $\mathbf{U} \in \mathbb{R}^{d_1 \times r}$  and  $\mathbf{V} \in \mathbb{R}^{d_2 \times r}$  are orthogonal matrices and  $\mathbf{S} \in \mathbb{R}^{r \times r}$  is a diagonal matrix of singular values. Then, the proximal operator of function  $\|\cdot\|_*$  at point  $\mathbf{T}_0$  with parameter  $\tau$  can be computed as follows:

$$\begin{aligned} \text{Prox}_{\tau}^{\|\cdot\|_*}(\mathbf{T}_0) &:= \underset{\mathbf{T}}{\text{argmin}} \frac{1}{2}\|\mathbf{T}_0 - \mathbf{T}\|_{\text{F}}^2 + \tau\|\mathbf{T}\|_* \\ &= \mathbf{U} \max(\mathbf{S} - \tau, 0)\mathbf{V}^{\text{T}}. \end{aligned}$$

Update the dual variables  $(\mathcal{A}, \{\mathcal{B}^k\})$ . We use dual ascending [33] to update  $(\mathcal{A}, \{\mathcal{B}^k\})$  as follows:

$$\begin{aligned} \mathcal{A}^{t+1} &= \mathcal{A}^t + \rho(\mathcal{K}^{t+1} - \mathcal{L}^{t+1}), \\ (\mathcal{B}^k)^{t+1} &= (\mathcal{B}^k)^t + \rho((\mathcal{T}^k)^{t+1} - \mathcal{L}^{t+1}), \quad k = 1, 2, 3. \end{aligned} \quad (60)$$

Termination Condition. Given a tolerance  $\epsilon > 0$ , check the termination condition of primal variables

$$\|\mathcal{X}^t - \mathcal{X}^t\|_{\infty} \leq \epsilon, \quad \forall \mathcal{X} \in \{\mathcal{L}, \mathcal{K}, \{\mathcal{T}^k\}\}, \quad (61)$$

and convergence of constraints

$$\|\mathcal{K}^t - \mathcal{L}^t\|_{\infty} \leq \epsilon, \quad \text{and} \quad \|(\mathcal{T}^k)^t - \mathcal{L}^t\|_{\infty} \leq \epsilon, \quad k = 1, 2, 3. \quad (62)$$

The ADMM-based algorithm is described in **Algorithm 1**.

**Algorithm 1.** ADMM for Problem(55)

---

**Require:**  $\{\mathcal{X}_i\}_i, \mathbf{y}, \rho, \epsilon, T_{\text{max}}$ .  
**Ensure:**  $\mathcal{L} = \mathcal{L}^t$ .  
1:  $\mathcal{L}^0 = \mathcal{K}^0 = (\mathcal{T}^1)^0 = (\mathcal{T}^2)^0 = (\mathcal{T}^3)^0 = \mathbf{0}, \mathcal{A}^0 = (\mathcal{B}^1)^0 = (\mathcal{B}^2)^0 = (\mathcal{B}^3)^0 = \mathbf{0}$ ;  
2: **for**  $t = 0$  to  $T_{\text{max}} - 1$  **do**  
3:   Update  $\mathcal{L}$  according to Eq. (57);  
4:   Update  $\mathcal{K}$  and  $\mathcal{T}^k$  according to Eq. (58) and Eq. (59), respectively;  
5:   Update the dual variables  $(\mathcal{A}, \{\mathcal{B}^k\})$  using Eq. (60);  
6:   Check stopping condition according to Eqs. (61) and (62);  
7: **end for**

---

Computational complexity analysis: We analyze the computational complexity as follows.

- By precomputing  $(\mathbb{I} + \mathfrak{X}^* \mathfrak{X} \mathfrak{X}^* \mathfrak{X})^{-1}$  and  $\mathfrak{X}^* \mathfrak{X}$ , which costs  $O(d_1^3 d_2^3 d_3^3 + Nd_1^2 d_2^2 d_3^2)$ , the cost of updating  $\mathcal{L}$  is  $O(d_1^2 d_2^2 d_3^2)$ .
- Updating  $\mathcal{K}$  and  $\mathcal{T}^k$  involves computing the proximal operator of TNN and NN, which costs  $O(d_1 d_2 d_3 (d_1 \wedge d_2 + \log d_3 + \sum_{k=1}^3 d_k \wedge d_{(k)}))$ .
- Updating  $\mathcal{A}$  and  $\{\mathcal{B}^k\}$  ( $k = 1, 2, 3$ ) costs  $O(d_1 d_2 d_3)$ .

Overall, supposing the iteration number is  $T$ , the total computational complexity will be

$$O\left(d_1^3 d_2^3 d_3^3 + T d_1^2 d_2^2 d_3^2 + T d_1 d_2 d_3 \left(d_1 \wedge d_2 + \log d_3 + \sum_{k=1}^3 d_k \wedge d_{(k)}\right)\right), \quad (63)$$

which is very expensive for large tensors. In some special cases (like tensor completion) where  $\langle \mathcal{X}_i, \mathcal{L} \rangle$  operates on an element of  $\mathcal{L}$ ,  $(\mathbb{I} + \mathfrak{X}^* \mathfrak{X} \mathfrak{X}^* \mathfrak{X})^{-1}$  and  $\mathfrak{X}^* \mathfrak{X}$  can be computed in  $O(d_1 d_2 d_3)$ . Hence, the total complexity of **Algorithm 1** will drop to

$$O\left(T d_1 d_2 d_3 \left(\min\{d_1, d_2\} + \log d_3 + \sum_{k=1}^3 d_k \wedge d_{(k)}\right)\right). \quad (64)$$

Convergence analysis: We then discuss the convergence of **Algorithm 1** as follows.

Theorem 4 (convergence of **Algorithm 1**). For any positive constant  $\rho$ , if the unaugmented Lagrangian function  $L_0(\mathcal{L}, \mathcal{K}, \{\mathcal{T}^k\}, \mathcal{A}, \{\mathcal{B}^k\})$  has a saddle point, then the iterations  $L_{\rho}(\mathcal{L}^t, \mathcal{K}^t, \{(\mathcal{T}^k)^t\}, \mathcal{A}^t, \{(\mathcal{B}^k)^t\})$  in **Algorithm 1** satisfy the residual convergence, objective convergence, and dual variable convergence (defined in [33]) of Problem (55) as  $t \rightarrow \infty$ .

Proof of Theorem 4. The key idea is to rewrite Problem(55) into a standard two-block ADMM problem. For notational simplicity, let

$$f(\mathbf{u}) = \frac{1}{2}\|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2^2, \quad g(\mathbf{v}) = \lambda\gamma\|\mathcal{K}\|_{\text{ttn}} + \lambda(1 - \gamma) \sum_{k=1}^3 \alpha_k \|\mathbf{T}_{(k)}^k\|_*,$$

with  $\mathbf{u}, \mathbf{v}, \mathbf{w}$ , and  $\mathbf{A}$  defined as follows:

$$\begin{aligned} \mathbf{u} &:= \text{vec}(\mathcal{L}) \in \mathbb{R}^{d_1 d_2 d_3}, & \mathbf{v} &:= \begin{bmatrix} \text{vec}(\mathcal{K}) \\ \text{vec}(\mathcal{T}^1) \\ \text{vec}(\mathcal{T}^2) \\ \text{vec}(\mathcal{T}^3) \end{bmatrix} \in \mathbb{R}^{4d_1 d_2 d_3}, \\ \mathbf{w} &:= \begin{bmatrix} \text{vec}(\mathcal{A}) \\ \text{vec}(\mathcal{B}^1) \\ \text{vec}(\mathcal{B}^2) \\ \text{vec}(\mathcal{B}^3) \end{bmatrix} \in \mathbb{R}^{4d_1 d_2 d_3}, & \mathbf{A} &:= \begin{bmatrix} \mathbf{I}_D \\ \mathbf{I}_D \\ \mathbf{I}_D \\ \mathbf{I}_D \end{bmatrix} \in \mathbb{R}^{4d_1 d_2 d_3 \times d_1 d_2 d_3}, \end{aligned}$$

where  $\text{vec}(\cdot)$  denotes the operation of tensor vectorization (see [18]).

It can be verified that  $f(\cdot)$  and  $g(\cdot)$  are closed, proper convex functions. Then, Problem(55) can be re-written as follows:

$$\begin{aligned} \min_{\mathbf{u}, \mathbf{v}} & f(\mathbf{u}) + g(\mathbf{v}) \\ \text{s.t.} & \mathbf{A}\mathbf{u} - \mathbf{v} = \mathbf{0}. \end{aligned}$$

According to the convergence analysis in [33], we have

$$\begin{aligned} \text{objective convergence :} & \lim_{t \rightarrow \infty} f(\mathbf{u}^t) + g(\mathbf{v}^t) = f^* + g^*, \\ \text{dual variable convergence :} & \lim_{t \rightarrow \infty} \mathbf{w}^t = \mathbf{w}^*, \\ \text{constraint convergence :} & \lim_{t \rightarrow \infty} \mathbf{A}\mathbf{u}^t - \mathbf{v}^t = \mathbf{0}, \end{aligned}$$

where  $f^*, g^*$  are the optimal values of  $f(\mathbf{u}), g(\mathbf{v})$ , respectively. Variable  $\mathbf{w}^*$  is a dual optimal point defined as

$$\mathbf{w}^* = \begin{bmatrix} \text{vec}(\mathcal{A}^*) \\ \text{vec}(\mathcal{B}^{1*}) \\ \text{vec}(\mathcal{B}^{2*}) \\ \text{vec}(\mathcal{B}^{3*}) \end{bmatrix},$$

where  $(\mathcal{A}^*, \{\mathcal{B}^{k*}\}_k)$  are the dual variables in a saddle point  $(\mathcal{L}^*, \mathcal{K}^*, \{\mathcal{T}^k\}^*, \mathcal{A}^*, \{\mathcal{B}^k\}^*)$  of the unaugmented Lagrangian  $L_0(\mathcal{L}, \mathcal{K}, \{\mathcal{T}^k\}, \mathcal{A}, \{\mathcal{B}^k\})$ .

Since there are only equality constraints in the convex problem (55), strong duality holds naturally as a corollary of Slater's condition [23], which further indicates that the unaugmented Lagrangian  $L_0(\mathcal{L}, \mathcal{K}, \{\mathcal{T}^k\}, \mathcal{A}, \{\mathcal{B}^k\})$  has a saddle point. Moreover, according to the analysis in [36], the convergence rate of general ADMM-based algorithms is  $O(1/T)$ , where  $T$  denotes the iteration number. In this way, the convergence behavior of **Algorithm 1** is analyzed.

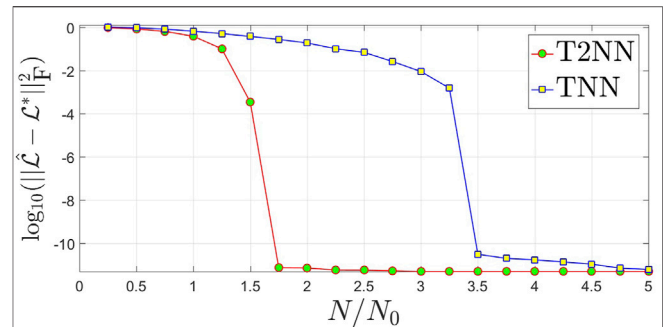
## 6 EXPERIMENTAL RESULTS

In this section, we first conduct experiments on synthetic datasets to validate the theory for tensor compressed sensing and then evaluate the effectiveness of the proposed T2NN on three types of real data for noisy tensor completion. MATLAB implementations of the algorithms are deployed on a PC running UOS system with an AMD 3 GHz CPU and a RAM of 40 GB.

### 6.1 Tensor Compressed Sensing

Our theoretical results on tensor compressed sensing are validated on synthetic data in this subsection. Motivated by [7], we consider a constrained T2NN minimization model that is equivalent to Model (15) for the ease of parameter selection. For performance evaluation, the proposed T2NN is also compared with TNN-based tensor compressed sensing [37]. First, the underlying tensor  $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  and its compressed observations  $\{y_i\}$  are synthesized by the following two steps, respectively:

- **Step 1: Generate  $\mathcal{L}^*$  that is low-rank in both spectral and original domains.** Given positive integers  $d_1, d_2, d_3$ , and  $r \leq \min\{d_1, d_2, d_3\}$ , we first generate  $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times r}$  by  $\mathcal{T} = \mathcal{G}_1 * \mathcal{G}_2$ , where  $\mathcal{G}_1 \in \mathbb{R}^{d_1 \times 1 \times r}$  and  $\mathcal{G}_2 \in \mathbb{R}^{1 \times d_2 \times r}$  are tensors with *i. i. d.* standard Gaussian entries. Then, let  $\mathcal{L}^* = \mathcal{T} \times_3 \mathbf{G}$  where  $\times_3$  is the tensor mode-3 product [18], and  $\mathbf{G} \in \mathbb{R}^{r \times d_3}$  is a matrix with *i. i. d.* standard Gaussian entries. Our extensive numerical experimental results show that with high probability, the tubal rank and Tucker rank of  $\mathcal{L}^*$  are all equal to  $r$ , that is,  $\text{rank}_{\text{tb}}(\mathcal{L}^*) = r$  and  $\text{rank}(\mathbf{L}_{(k)}^*) = r, \forall k = 1, 2, 3$ .
- **Step 2: Generate  $N$  compressed observations  $\{y_i\}$ .** Given a positive integer  $N \ll D$ , we first generate  $N$  design tensors  $\{\mathcal{X}_i\}$  with *i. i. d.* standard Gaussian entries. Then,  $N$  noise variables  $\{\xi_i\}$  are generated as *i. i. d.* standard Gaussian variables. The parameter of standard deviation  $\sigma$  is set by  $\sigma = c\sigma_0$ , where  $\sigma_0 = \|\mathcal{L}^*\|_{\text{F}} / \sqrt{d_1 d_2 d_3}$ , and we use  $c$  to denote the noise level. Finally,  $\{y_i\}$  are formed according to the



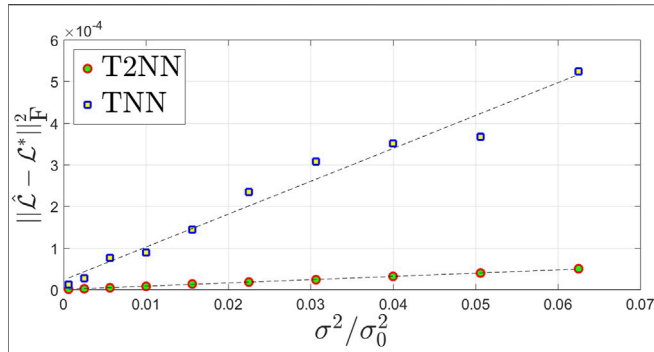
**FIGURE 1** | Estimation error in logarithm vs. the normalized observation number  $N/N_0$  for tensor compressed sensing of underlying tensors of size  $16 \times 16 \times 16$  and rank proxy  $r = 2$ . The proposed T2NN is compared with TNN [37].

observation model (13). The goal of tensor compressed sensing is to reconstruct the known  $\mathcal{L}^*$  from its noisy compressed observations  $\{y_i\}$ .

For simplicity, we consider cubic tensors, i.e.,  $d_1 = d_2 = d_3 = d$ , and choose the parameter of T2NN by  $\gamma = 1/4, \alpha_1 = \alpha_2 = \alpha_3 = 1/3$ . Recall that the underlying tensor  $\mathcal{L}^* \in \mathbb{R}^{d \times d \times d}$  generated by the above Step 1 has the tubal rank and Tucker rank all equal to  $r$  with high probability. We consider tensors with dimensionality  $d \in \{16, 20, 24\}$  and rank proxy  $r \in \{2, 3\}$ . Then, if the proposed main theorem for tensor compressed sensing (i.e., Theorem 2) is correct, the following two phenomena should be observed:

- (1) **Phenomenon 1:** In the noiseless setting, i.e.,  $\sigma = 0$ , if the observation number  $N$  is larger than  $C_0 r d^2$  for a sufficiently large constant  $C_0$ , then the estimation error  $\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2$  can be zero, which means exact recovery. Let  $N_0 = r d^2$  as a unit measure of the sample complexity. Then, by increasing the observation number  $N$  gradually from 0, we will observe a phase transition point of the estimation error in the noiseless setting: If  $N/N_0 > C_0$ , the estimation error is relatively “large”; once  $N/N_0 \leq C_0$ , the error will drop dramatically to 0.
- (2) **Phenomenon 2:** In the noisy case, the estimation error  $\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2$  scales linearly with the variance  $\sigma^2$  of the random noises once the observation number  $N \geq C_0 N_0$ .

To check whether Phenomenon 1 occurs, we conduct tensor compressed sensing by setting the noise variance  $\sigma^2 = 0$ . We gradually increase the normalized observation number  $N/N_0$  from 0.25 to 5. For each different setting of  $d, r$ , and  $N/N_0$ , we repeat the experiments 10 times and report the averaged estimation error  $\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2$ . For both TNN [37] and the proposed T2NN, we plot the curves of estimation error in logarithm versus the normalized observation number  $N/N_0$  for  $\mathcal{L}^* \in \mathbb{R}^{d \times d \times d}$  with rank proxy  $r = 2$  in **Figure 1**. It can be seen that Phenomenon 1 occurs for the proposed T2NN: When  $N/N_0 > 1.75$ , the estimation error is relatively “large”; once  $N/N_0 \leq 1.75$ , the error will drop dramatically to 0. The same phenomenon also occurs for TNN with a phase transition point near 3.5. Thus, the sample complexity for exact tensor compressed sensing of T2NN



**FIGURE 2** | Estimation error vs. the (squared) noise level  $\sigma^2/\sigma_0^2$  for tensor compressed sensing of underlying tensors of size  $16 \times 16 \times 16$  and rank proxy  $r = 2$ . The proposed T2NN is compared with TNN [37].

is lower than that of TNN, indicating the superiority of the proposed T2NN. Since similar phenomena have also been observed for tensors of other sizes and rank proxies, we simply omit them.

For the validation of Phenomenon 2, we consider the noisy settings with normalized sample complexity  $N/N_0 = 3.5$ , which is nearly the phase transition point of TNN and much greater than that of T2NN. We gradually increase the noise level  $c = \sigma/\sigma_0$  from 0.025 to 0.25. For each different setting of  $d$ ,  $r$ , and  $c$ , we repeat the experiments 10 times and report the averaged estimation error  $\|\hat{\mathcal{L}} - \mathcal{L}^*\|_F^2$ . For both TNN [37] and the proposed T2NN, we plot the curves of estimation error in logarithm versus the (squared) noise level  $\sigma^2/\sigma_0^2$  for  $\mathcal{L}^* \in \mathbb{R}^{d \times d \times d}$  with rank proxy  $r = 2$  in **Figure 2**. It can be seen that Phenomenon 2 also occurs for the proposed T2NN: The estimation error scales approximately linearly with the (squared) noise level. The same phenomenon can also be observed for TNN with a higher estimation error than T2NN, indicating T2NN is more accurate than TNN. We omit the results for tensors of other sizes and rank proxies because the error curves are so similar to **Figure 2**.

## 6.2 Noisy Tensor Completion

This subsection evaluates effectiveness of the proposed T2NN through performance comparison with matrix nuclear norms (NN) [30], SNN [22], and TNN [25] by carrying out noisy tensor completion on three different types of visual data including video data, hyperspectral images, and seismic data.

### 6.2.1 Experimental Settings

Given the tensor data  $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ , the goal is to recover it from its partial noisy observations. We consider uniform sampling with ratio  $p \in \{0.05, 0.1, 0.15\}$  for the tensors, that is,  $\{95, 90, 85\}\%$  entries of a tensor are missing. The noise follows *i. i. d.* Gaussian  $\mathcal{N}(0, \sigma^2)$  where  $\sigma = 0.05\sigma_0$ , where  $\sigma_0 = \|\mathcal{L}^*\|_F / \sqrt{d_1 d_2 d_3}$  is the rescaled magnitude of tensor  $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ .

**TABLE 1** | PSNR and SSIM values obtained by four norms (NN [30], SNN [22], TNN [25], and our T2NN) for noisy tensor completion on the YUV videos.

#### (a) Akiyo

Sampling ratio	Index	NN	SNN	TNN	T2NN
5%	PSNR	15.37	19.49	27.04	<b>27.51</b>
	SSIM	0.1864	0.6047	0.8019	<b>0.8302</b>
10%	PSNR	18.01	22.54	29.18	<b>30.08</b>
	SSIM	0.2858	0.7186	0.8556	<b>0.8828</b>
15%	PSNR	19.64	24.37	30.60	<b>31.43</b>
	SSIM	0.3694	0.7812	0.8791	<b>0.8968</b>

#### (b) Carphone

Sampling ratio	Index	NN	SNN	TNN	T2NN
5%	PSNR	13.58	17.58	23.81	<b>24.16</b>
	SSIM	0.1378	0.5282	0.6725	<b>0.7226</b>
10%	PSNR	16.04	20.49	25.38	<b>25.87</b>
	SSIM	0.2352	0.6425	0.7309	<b>0.7813</b>
15%	PSNR	17.86	22.42	26.40	<b>26.81</b>
	SSIM	0.3242	0.7139	0.7663	<b>0.8057</b>

#### (c) Grandma

Sampling ratio	Index	NN	SNN	TNN	T2NN
5%	PSNR	16.52	19.13	28.53	<b>28.84</b>
	SSIM	0.1928	0.5503	0.8135	<b>0.8516</b>
10%	PSNR	18.34	22.52	31.44	<b>32.64</b>
	SSIM	0.2992	0.6755	0.8822	<b>0.9141</b>
15%	PSNR	19.66	24.88	33.02	<b>34.28</b>
	SSIM	0.3757	0.7525	0.9088	<b>0.9307</b>

#### (d) Mother-daughter

Sampling ratio	Index	NN	SNN	TNN	T2NN
5%	PSNR	16.17	20.22	27.66	<b>27.96</b>
	SSIM	0.1895	0.5824	0.7491	<b>0.7816</b>
10%	PSNR	18.55	23.48	29.38	<b>30.32</b>
	SSIM	0.2780	0.6840	0.8035	<b>0.8412</b>
15%	PSNR	20.32	25.31	30.49	<b>31.18</b>
	SSIM	0.3548	0.7405	0.8293	<b>0.8526</b>

The highest PSNR/SSIM are highlighted in bold.

### 6.2.2 Performance evaluation

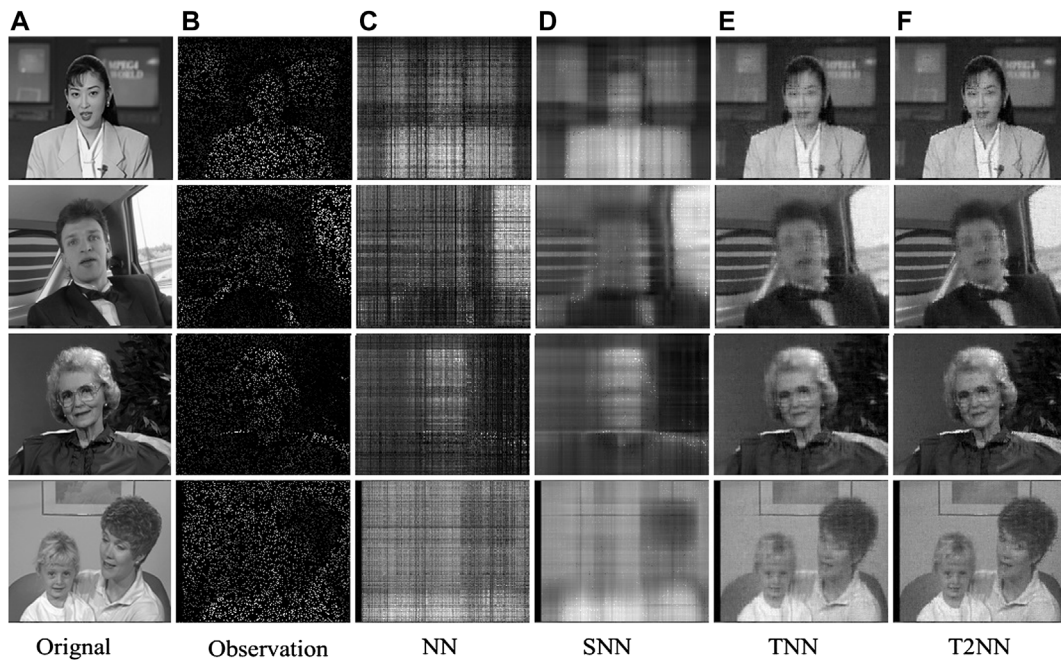
The effectiveness of algorithms is measured by the Peak Signal Noise Ratio (PSNR) and structural similarity (SSIM) [38]. Specifically, the PSNR of an estimator  $\hat{\mathcal{L}}$  is defined as

$$\text{PSNR} := 10 \log_{10} \left( \frac{d_1 d_2 d_3 \|\mathcal{L}^*\|_\infty^2}{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_F^2} \right),$$

for the underlying tensor  $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ . The SSIM is computed via

$$\text{SSIM} := \frac{(2\mu_{\mathcal{L}^*} \mu_{\hat{\mathcal{L}}} + (0.01\bar{\omega})^2)(2\sigma_{\mathcal{L}^*, \hat{\mathcal{L}}} + (0.03\bar{\omega})^2)}{(\mu_{\mathcal{L}^*}^2 + \mu_{\hat{\mathcal{L}}}^2 + (0.01\bar{\omega})^2)(\sigma_{\mathcal{L}^*}^2 + \sigma_{\hat{\mathcal{L}}}^2 + (0.03\bar{\omega})^2)},$$

where  $\mu_{\mathcal{L}^*}, \mu_{\hat{\mathcal{L}}}, \sigma_{\mathcal{L}^*}, \sigma_{\hat{\mathcal{L}}}, \sigma_{\mathcal{L}^*, \hat{\mathcal{L}}}$ , and  $\bar{\omega}$  denote the local means, standard deviation, cross-covariance, and dynamic range of the magnitude of tensors  $\mathcal{L}^*$  and  $\hat{\mathcal{L}}$ . Larger PSNR and SSIM values indicate the higher quality of the estimator  $\hat{\mathcal{L}}$ . In each setting, we test



**FIGURE 3** | Visual results obtained by four norms for noisy tensor completion with 95% missing entries on the YUV-video dataset. The first to fourth rows correspond to the video of Akiyo, Carphone, Grandman, and Mother-daughter, respectively. The sub-plots from (A) to (F): (A) a frame of the original video, (B) the observed frame, (C) the frame recovered by NN [30], (D) the frame recovered by SNN [22], (E) the frame recovered by the vanilla TNN [25], and (F) the frame recovered by our T2NN.

each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values.

### 6.2.3 Parameter Setting

For NN [30], we set the parameter  $\lambda = \lambda_i \sigma \sqrt{p(d_1 \vee d_2) \log(d_1 + d_2)}$ . For SNN [22], we set the regularization parameter  $\lambda = \lambda_i$  and chose the weight  $\alpha$  by  $\alpha_1: \alpha_2: \alpha_3 = 1:1:1$ . For TNN [25], we set  $\lambda = \lambda_i \sigma \sqrt{pd_3(d_1 \vee d_2) \log(d_1 d_3 + d_2 d_3)}$ . For the proposed T2NN, we set the regularization parameter  $\lambda = \lambda_i \sigma \sqrt{pd_3(d_1 \vee d_2) \log(d_1 d_3 + d_2 d_3)}$  and choose the weights  $\gamma = 0.5$  and  $\alpha$  with  $\alpha_1: \alpha_2: \alpha_3 = 1:1:10$ . The factor  $\lambda_i$  is then tuned in  $\{10^{-3}, 10^{-2}, \dots, 10^3\}$  for each norm, and we chose the one with highest PSNRs in most cases in the parameter tuning phase.

### 6.2.4 Experiments on Video Data

We first conduct noisy video completion on four widely used YUV videos: *Akiyo*, *Carphone*, *Grandma*, and *Mother-daughter*. Owing to computational limitation, we simply use the first 30 frames of the Y components of all the videos and obtain four tensors of size  $144 \times 17 \times 30$ . We first report the averaged PSNR and SSIM values obtained by four norms for quantitative comparison in Table 1 and then give visual examples in Figure 3 when 95% of the tensor entries are missing for qualitative evaluation. A demo of the source code is available at <https://github.com/pingzaiwang/T2NN-demo>.

**TABLE 2** | PSNR and SSIM values obtained by four norms (NN [30], SNN [22], TNN [25], and our T2NN) for noisy tensor completion on the hyperspectral datasets.

#### (a) Indian pines

Sampling ratio	Index	NN	SNN	TNN	T2NN
5%	PSNR	20.44	22.01	25.68	<b>26.00</b>
	SSIM	0.3895	0.6359	0.6293	<b>0.6730</b>
10%	PSNR	22.23	24.94	27.45	<b>28.27</b>
	SSIM	0.4836	0.7171	0.7226	<b>0.7724</b>
15%	PSNR	23.52	26.61	28.54	<b>29.11</b>
	SSIM	0.5438	0.7668	0.7713	<b>0.7979</b>

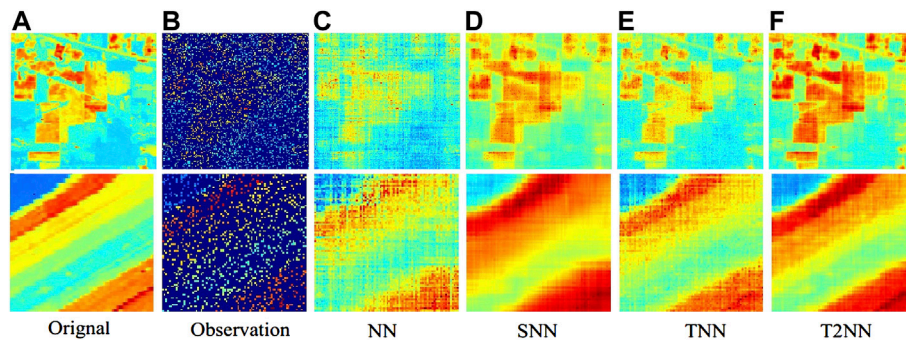
#### (b) Salinas A

Sampling ratio	Index	NN	SNN	TNN	T2NN
5%	PSNR	15.21	20.79	22.55	<b>23.68</b>
	SSIM	0.2594	<b>0.7547</b>	0.5667	0.7013
10%	PSNR	20.62	25.56	25.72	<b>27.93</b>
	SSIM	0.4775	0.8284	0.7027	<b>0.8291</b>
15%	PSNR	23.09	27.99	28.06	<b>29.67</b>
	SSIM	0.5643	0.8622	0.7804	<b>0.8671</b>

The highest PSNRs/SSIM are highlighted in bold.

### 6.2.5 Experiments on Hyperspectral Data

We then carry out noisy tensor completion on subsets of the two representative hyperspectral datasets described as follows:



**FIGURE 4** | Visual results obtained by four norms for noisy tensor completion with 85% missing entries on the hyperspectral dataset (gray data shown with pseudo-color). The first and second rows correspond to *Indian Pines* and *Salinas A*, respectively. The sub-plots from (A) to (F): (A) a frame of the original data, (B) the observed frame, (C) the frame recovered by NN [30], (D) the frame recovered by SNN [22], (E) the frame recovered by the vanilla TNN [25], and (F) the frame recovered by our T2NN.

**TABLE 3** | PSNR and SSIM values obtained by four norms (NN [30], SNN [22], TNN [25], and our T2NN) for noisy tensor completion on the *Seismic* dataset.

Sampling ratio (%)	Index	NN	SNN	TNN	T2NN
5	PSNR	22.25	22.28	22.49	<b>22.80</b>
	SSIM	0.4369	<b>0.4906</b>	0.3928	0.4794
10	PSNR	23.58	23.53	23.48	<b>24.05</b>
	SSIM	0.5462	0.5740	0.5004	<b>0.5845</b>
15	PSNR	24.74	24.66	24.51	<b>25.25</b>
	SSIM	0.6266	0.6552	0.5898	<b>0.6657</b>

The highest PSNR/SSIM are highlighted in bold.

- *Indian Pines*: The dataset was collected by AVIRIS sensor in 1992 over the Indian Pines test site in North-western Indiana and consists of  $145 \times 145$  pixels and 224 spectral reflectance bands. We use the first 30 bands in the experiments due to the trade-off between the limitation of computing resources.
- *Salinas A*: The data were acquired by AVIRIS sensor over the Salinas Valley, California in 1998, and consists of 224 bands over a spectrum range of 400–2500 nm. This dataset has a spatial extent of  $86 \times 83$  pixels with a resolution of 3.7 m. We use the first 30 bands in the experiments too.

The averaged PSNR and SSIM values are given in Table 2 for quantitative comparison. We also show visual examples in Figure 4 when 85% of the tensor entries are missing for qualitative evaluation.

### 6.2.6 Experiments on Seismic Data

We use the seismic data tensor of size  $512 \times 512 \times 3$ , which is abstracted from the test data “seismic.mat” of a toolbox for seismic data processing from Center of Geophysics, Harbin Institute of Technology, China. For quantitative comparison, we present the PSNR and SSIM values for two sampling schemes in Table 3.

### 6.2.7 Summary and Analysis of Experimental Results

According to the experimental results on three types of real tensor data shown in Table 1, Table 2, Table 3, and Figure 3, the summary and analysis are presented as follows:

- 1) In all the cases, tensor norms (SNN, TNN, and T2NN) perform better than the matrix norm (NN). It can be explained that tensor norms can honestly preserve the multi-way structure of tensor data such that the rich inter-modal and intra-modal correlations of the data can be exploited to impute the missing values, whereas the matrix norm can only handle two-way structure and thus fails to model the multi-way structural correlations of the tensor data.
- 2) In most cases, TNN outperforms SNN, which is in consistency with the results reported in [14, 17, 25]. One explanation is that the used video, hyperspectral images, and seismic data all possess stronger low-rankness in the spectral domain (than in the original domain), which can be successfully captured by TNN.
- 3) In most cases, the proposed T2NN performs best among the four norms. We owe the promising performance to the capability of T2NN in simultaneously exploiting low-rankness in both spectral and original domains.

## 7 CONCLUSION AND DISCUSSIONS

### 7.1 Conclusion

Due to its definition solely in the spectral domain, the popular TNN may be incapable to exploit low-rankness in the original domain. To remedy this weaknesses, a hybrid tensor norm named the “Tubal + Tucker” Nuclear Norm (T2NN) was first defined as the weighted sum of TNN and SNN to model both spectral and original domain low-rankness. It was further used to formulate a penalized least squares estimator for tensor recovery from noisy linear observations. Upper bounds on the estimation error were established in both deterministic and non-asymptotic senses to analyze the statistical performance of



the proposed estimator. An ADMM-based algorithm was also developed to efficiently compute the estimator. The effectiveness of the proposed model was demonstrated through experimental results on both synthetic and real datasets.

## 7.2 Limitations of the Proposed Model and Possible Solutions

Generally speaking, the proposed estimator has the following two drawbacks due to the adoption of T2NN:

- *Sample inefficiency:* The analysis of [24, 28] indicates that for tensor recovery from a small number of observations, T2NN cannot provide essentially lower sample complexity than TNN.
- *Computational inefficiency:* Compared to TNN, T2NN is more time-consuming since it involves computing both TNN and SNN.

We list several directions that this work can be extended to overcome the above drawbacks.

- *For sample inefficiency:* First, inspired by the attempt of adopting the “best” norm (e.g., Eq. 8 in [28]), the following model can be considered:

$$\min_{\mathcal{L}} \max \left\{ \frac{\|\mathcal{L}\|_{\text{tnn}}}{\|\mathcal{L}^*\|_{\text{tnn}}}, \max_{k=1,2,3} \frac{\|\mathbf{L}_{(k)}\|_*}{\|\mathbf{L}_{(k)}^*\|_*} \right\} \quad (65)$$

s.t.  $\|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2 \leq \epsilon$

for a certain noise level  $\epsilon \geq 0$ . Although Model (65) has a significantly higher accuracy and lower sample complexity according to the analysis in [28], it is impractical because it requires  $\|\mathcal{L}^*\|_{\text{tnn}}$  and  $\|\mathbf{L}_{(k)}^*\|_*$  ( $k = 1, 2, 3$ ), which are unknown in advance. Motivated by [39], a more practical model is given as follows:

$$\min_{\mathcal{L}} \sum_{k=1}^3 \exp(\alpha_k \|\mathbf{L}_{(k)}\|_*) + \exp(\beta \|\mathcal{L}\|_{\text{tnn}})$$

s.t.  $\|\mathbf{y} - \mathfrak{X}(\mathcal{L})\|_2 \leq \epsilon$ ,

where  $\beta > 0$  is a regularization parameter.

- *For computational inefficiency:* To improve the efficiency of the proposed T2NN-based models, we can use more efficient solvers of Problem (15) by adopting the factorization strategy [40, 41] or sampling-based approaches [42].

## 7.3 Extensions to the Proposed Model

In this subsection, we discuss possible extensions of the proposed model to general  $K$ -order ( $K > 3$ ) tensors, general spectral domains, robust tensor recovery, and multi-view learning, respectively.

- *Extensions to  $K$ -order ( $K > 3$ ) tensors:* Currently, the proposed T2NN is defined solely for 3-order tensors, and it cannot be directly applied to tensors of more than

3 orders like color videos. For general  $K$ -order tensors, it is suggested to replace the tubal nuclear norm in the definition of T2NN with orientation invariant tubal nuclear norm [5], which is defined to exploit multi-orientational spectral low-rankness for general higher-order tensors.

- *Extensions to general spectral and original domains:* This paper considers the DFT-based tensor product for spectral low-rank modeling. Recently, the DFT based t-product has been generalized to the  $^*_L$ -product defined via any invertible linear transform [43], under which the tubal nuclear norm is also extended to  $^*_L$ -tubal nuclear norm [44] and  $^*_L$ -Spectral  $k$ -support norm [7]. It is natural to generalize the proposed T2NN by changing the tubal nuclear norm to  $^*_L$ -tubal nuclear norm or  $^*_L$ -Spectral  $k$ -support norm for further extensions. It is also interesting to consider other tensor decompositions for original domain low-rankness modeling such as CP, TT, and TR as future work.
- *Extensions to robust tensor recovery:* In many real applications, the tensor signal may also be corrupted by gross sparse outliers. Motivated by [5], the proposed T2NN can also be used in resisting sparse outliers for robust tensor recovery as follows:

$$\min_{\mathcal{L}, \mathcal{S}} \frac{1}{2} \|\mathbf{y} - \mathfrak{X}(\mathcal{L} + \mathcal{S})\|_2 + \lambda \|\mathcal{L}\|_{\text{t2nn}} + \mu \|\mathcal{S}\|_1,$$

where  $\mathcal{S} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$  denotes the tensor of sparse outliers, the tensor  $l_1$ -norm  $\|\cdot\|_1$  is applied to encourage sparsity in  $\mathcal{S}$ , and  $\mu > 0$  is a regularization parameter.

- *Extensions to multi-view learning:* Due to its superiority in modeling multi-linear correlations of multi-modal data, TNN has been successfully applied to multi-view self-representations for clustering [45, 46]. Our proposed T2NN can also be utilized for clustering by straightforwardly replacing TNN in the formulation of multi-view learning models (e.g., Eq. 9 in [45]).

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here; <https://sites.google.com/site/subudhibadri/fewhelpfuldownloads>, <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>, [https://rslab.ut.ac.ir/documents/81960329/82035173/SalinasA\\_corrected.mat](https://rslab.ut.ac.ir/documents/81960329/82035173/SalinasA_corrected.mat), <https://github.com/sevenysw/MathGeo2018>.

## AUTHOR CONTRIBUTIONS

Conceptualization and methodology—YL and AW; software—AW; formal analysis—YL, AW, GZ, and QZ; resources—YL, GZ, and QZ; writing: original draft preparation—YL, AW, GZ, and QZ; project

administration and supervision—GZ, and QZ; and funding acquisition—AW, GZ, and QZ. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants 61872188, 62073087, 62071132, 62103110, 61903095, U191140003, and 61973090, in part by the China Postdoctoral Science Foundation under Grant 2020M672536, and in part by the Natural Science

Foundation of Guangdong Province under Grants 2020A1515010671, 2019B010154002, and 2019B010118001.

## ACKNOWLEDGMENTS

AW is grateful to Prof. Zhong Jin in Nanjing University of Science and Technology for his long-time and generous support in both research and life. In addition, he would like to thank the Jin family in Zhuzhou for their kind understanding in finishing the project of tensor learning in these years.

## REFERENCES

- Guo C, Modi K, Poletti D. Tensor-Network-Based Machine Learning of Non-Markovian Quantum Processes. *Phys Rev A* (2020) 102:062414.
- Ma X, Zhang P, Zhang S, Duan N, Hou Y, Zhou M, et al. A Tensorized Transformer for Language Modeling. *Adv Neural Inf Process Syst* (2019) 32.
- Meng Y-M, Zhang J, Zhang P, Gao C, Ran S-J. Residual Matrix Product State for Machine Learning. *arXiv preprint arXiv:2012.11841* (2020).
- Ran S-J, Sun Z-Z, Fei S-M, Su G, Lewenstein M. Tensor Network Compressed Sensing with Unsupervised Machine Learning. *Phys Rev Res* (2020) 2:033293. doi:10.1103/physrevresearch.2.033293
- Wang A, Zhao Q, Jin Z, Li C, Zhou G. Robust Tensor Decomposition via Orientation Invariant Tubal Nuclear Norms. *Sci China Technol Sci* (2022) 34: 6102. doi:10.1007/s11431-021-1976-2
- Zhang X, Ng MK-P. Low Rank Tensor Completion with Poisson Observations. *IEEE Trans Pattern Anal Machine Intelligence* (2021). doi:10.1109/tpami.2021.3059299
- Wang A, Zhou G, Jin Z, Zhao Q. Tensor Recovery via  $\ell_1$ -Spectral  $k$ -Support Norm. *IEEE J Sel Top Signal Process* (2021) 15:522–34. doi:10.1109/jstsp.2021.3058763
- Cui C, Zhang Z. High-Dimensional Uncertainty Quantification of Electronic and Photonic Ic with Non-Gaussian Correlated Process Variations. *IEEE Trans Computer-Aided Des Integrated Circuits Syst* (2019) 39:1649–61. doi:10.1109/TCAD.2019.2925340
- Liu X-Y, Aeron S, Aggarwal V, Wang X. Low-Tubal-Rank Tensor Completion Using Alternating Minimization. *IEEE Trans Inform Theor* (2020) 66:1714–37. doi:10.1109/tit.2019.2959980
- Carroll JD, Chang J-J. Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of “Eckart-Young” Decomposition. *Psychometrika* (1970) 35:283–319. doi:10.1007/bf02310791
- Tucker LR. Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika* (1966) 31:279–311. doi:10.1007/bf02289464
- Oseledets IV. Tensor-Train Decomposition. *SIAM J Sci Comput* (2011) 33: 2295–317. doi:10.1137/090752286
- Zhao Q, Zhou G, Xie S, Zhang L, Cichocki A. Tensor Ring Decomposition. *arXiv preprint arXiv:1606.05535* (2016).
- Zhang Z, Ely G, Aeron S, Hao N, Kilmer M. Novel Methods for Multilinear Data Completion and De-Noiseing Based on Tensor-Svd. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014). p. 3842–9. doi:10.1109/cvpr.2014.485
- 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) 34:148–72. doi:10.1137/110837711
- Hou J, Zhang F, Qiu H, Wang J, Wang Y, Meng D. Robust Low-Tubal-Rank Tensor Recovery from Binary Measurements. *IEEE Trans Pattern Anal Machine Intelligence* (2021). doi:10.1109/tpami.2021.3063527
- 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) 42:925–38. doi:10.1109/tpami.2019.2891760
- Kolda TG, Bader BW. Tensor Decompositions and Applications. *SIAM Rev* (2009) 51:455–500. doi:10.1137/07070111x
- Li X, Wang A, Lu J, Tang Z. Statistical Performance of Convex Low-Rank and Sparse Tensor Recovery. *Pattern Recognition* (2019) 93:193–203. doi:10.1016/j.patcog.2019.03.014
- Liu J, Musialski P, Wonka P, Ye J. Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Trans Pattern Anal Mach Intell* (2013) 35:208–20. doi:10.1109/tpami.2012.39
- Qiu Y, Zhou G, Chen X, Zhang D, Zhao X, Zhao Q. Semi-Supervised Non-Negative Tucker Decomposition for Tensor Data Representation. *Sci China Technol Sci* (2021) 64:1881–92. doi:10.1007/s11431-020-1824-4
- Tomioka R, Suzuki T, Hayashi K, Kashima H. Statistical Performance of Convex Tensor Decomposition. In: Proceedings of Annual Conference on Neural Information Processing Systems (2011). p. 972–80.
- Boyd S, Boyd SP, Vandenberghe L. *Convex Optimization*. Cambridge: Cambridge University Press (2004).
- Mu C, Huang B, Wright J, Goldfarb D. Square Deal: Lower Bounds and Improved Relaxations for Tensor Recovery. In: International Conference on Machine Learning (2014). p. 73–81.
- Wang A, Lai Z, Jin Z. Noisy Low-Tubal-Rank Tensor Completion. *Neurocomputing* (2019) 330:267–79. doi:10.1016/j.neucom.2018.11.012
- Zhou P, Lu C, Lin Z, Zhang C. Tensor Factorization for Low-Rank Tensor Completion. *IEEE Trans Image Process* (2018) 27:1152–63. doi:10.1109/tip.2017.2762595
- Negahban S, Wainwright MJ. Estimation of (Near) Low-Rank Matrices with Noise and High-Dimensional Scaling. *Ann Stat* (2011) 2011:1069–97. doi:10.1214/10-aos850
- Oymak S, Jalali A, Fazel M, Eldar YC, Hassibi B. Simultaneously Structured Models with Application to Sparse and Low-Rank Matrices. *IEEE Trans Inform Theor* (2015) 61:2886–908. doi:10.1109/tit.2015.2401574
- Foucart S, Rauhut H. *A Mathematical Introduction to Compressive Sensing, Vol. 1*. Basel, Switzerland: Birkhäuser Basel (2013).
- Klopp O. Noisy Low-Rank Matrix Completion with General Sampling Distribution. *Bernoulli* (2014) 20:282–303. doi:10.3150/12-bej486
- Klopp O. Matrix Completion by Singular Value Thresholding: Sharp Bounds. *Electron J Stat* (2015) 9:2348–69. doi:10.1214/15-ejs1076
- Vershynin R. *High-Dimensional Probability: An Introduction with Applications in Data Science, Vol. 47*. Cambridge: Cambridge University Press (2018).
- Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations Trends® Machine Learn* (2011) 3:1–122. doi:10.1561/22000000016
- Wang A, Wei D, Wang B, Jin Z. Noisy Low-Tubal-Rank Tensor Completion Through Iterative Singular Tube Thresholding. *IEEE Access* (2018) 6: 35112–28. doi:10.1109/access.2018.2850324
- Cai J-F, Candès EJ, Shen Z. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM J Optim* (2010) 20:1956–82. doi:10.1137/080738970
- He B, Yuan X. On the  $\mathcal{O}(1/n)$  Convergence Rate of the Douglas-Rachford Alternating Direction Method. *SIAM J Numer Anal* (2012) 50:700–9. doi:10.1137/110836936

37. Lu C, Feng J, Lin Z, Yan S. Exact Low Tubal Rank Tensor Recovery from Gaussian Measurements. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (2018). p. 1948–54. doi:10.24963/ijcai.2018/347
38. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Trans Image Process* (2004) 13:600–12. doi:10.1109/tip.2003.819861
39. Zhang X, Zhou Z, Wang D, Ma Y. Hybrid Singular Value Thresholding for Tensor Completion. In: Twenty-Eighth AAAI Conference on Artificial Intelligence (2014). p. 1362–8.
40. Wang A-D, Jin Z, Yang J-Y. A Faster Tensor Robust Pca via Tensor Factorization. *Int J Mach Learn Cyber* (2020) 11:2771–91. doi:10.1007/s13042-020-01150-2
41. Liu G, Yan S. Active Subspace: Toward Scalable Low-Rank Learning. *Neural Comput* (2012) 24:3371–94. doi:10.1162/neco\_a\_00369
42. Wang L, Xie K, Semong T, Zhou H. Missing Data Recovery Based on Tensor-Cur Decomposition. *IEEE Access* (2017) PP:1.
43. Kernfeld E, Kilmer M, Aeron S. Tensor-Tensor Products with Invertible Linear Transforms. *Linear Algebra its Appl* (2015) 485:545–70. doi:10.1016/j.laa.2015.07.021
44. Lu C, Peng X, Wei Y. Low-Rank Tensor Completion with a New Tensor Nuclear Norm Induced by Invertible Linear Transforms. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019). p. 5996–6004. doi:10.1109/cvpr.2019.00615
45. Lu G-F, Zhao J. Latent Multi-View Self-Representations for Clustering via the Tensor Nuclear Norm. *Appl Intelligence* (2021) 2021:1–13. doi:10.1007/s10489-021-02710-x
46. Liu Y, Zhang X, Tang G, Wang D. Multi-View Subspace Clustering Based on Tensor Schatten-P Norm. In: 2019 IEEE International Conference on Big Data (Big Data). Los Angeles, CA, USA: IEEE (2019). p. 5048–55. doi:10.1109/bigdata47090.2019.9006347

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Luo, Wang, Zhou and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.