

Article

Stable Tensor Principal Component Pursuit: Error Bounds and Efficient Algorithms

Wei Fang ^{1,*}, Dongxu Wei ² and Ran Zhang ³

¹ Department of Computer Science and Technology, Huaibei Vocational and Technical College, Huaibei 235000, China

² School of Physics and Electronic Electrical Engineering, Huaiyin Normal University, Huaian 223300, China; weidx@hytc.edu.cn

³ Mathematics Teaching and Research Group, Nanjing No.9 High School, Nanjing 210018, China; exjzhang@163.com

* Correspondence: fangweixyz@sina.com

Received: 13 November 2019; Accepted: 29 November 2019; Published: 3 December 2019



Abstract: The rapid development of sensor technology gives rise to the emergence of huge amounts of tensor (i.e., multi-dimensional array) data. For various reasons such as sensor failures and communication loss, the tensor data may be corrupted by not only small noises but also gross corruptions. This paper studies the Stable Tensor Principal Component Pursuit (STPCP) which aims to recover a tensor from its corrupted observations. Specifically, we propose a STPCP model based on the recently proposed tubal nuclear norm (TNN) which has shown superior performance in comparison with other tensor nuclear norms. Theoretically, we rigorously prove that under tensor incoherence conditions, the underlying tensor and the sparse corruption tensor can be stably recovered. Algorithmically, we first develop an ADMM algorithm and then accelerate it by designing a new algorithm based on orthogonal tensor factorization. The superiority and efficiency of the proposed algorithms is demonstrated through experiments on both synthetic and real data sets.

Keywords: tensor principal component pursuit; stable recovery; tensor SVD; ADMM

1. Introduction

In recent years, different types of tensor data have emerged with the significant progress of modern sensor technology, such as color images [1], videos [2], functional MRI data [3], hyper-spectral images [4], point cloud data [5], traffic stream data [6], etc. Thanks to its multi-way nature, tensor-based methods have natural superiority over vector and matrix-based methods in analyzing and processing ubiquitous modern multi-way data, and have found extensive applications in computer vision [1,7], data mining [5], machine learning [2], signal processing [8], to name a few. In real applications, the acquired tensor data may often suffer from noises and gross corruptions owing to many different reasons such as sensor failure, lens pollution, communication interference, occlusion in videos, or abnormalities in a sensor network [9], etc. At the same time, many real-world tensor data, such as face images or videos, have been shown to have some low-dimensional structure and can be well approximated by a smaller number of “principal components” [8]. Then, a question naturally arises: how to pursue the principal components of an observed tensor data in the presence of both noises and gross corruptions? We will answer this question in this paper and refer to the proposed methodology as Stable Tensor Principal Component Pursuit (STPCP).

The tensor low-rankness is an ideal model of the property that a tensor data can be well approximated by a small number of principal components [8]. In the last decade, low-rank tensor

models have attracted much attention in many fields [10]. There are multiple low-rank tensor models since there exist different definitions of tensor rank. Among these models, the low CP rank model [11] and the low Tucker rank model [1] should be the most famous two. The low CP rank model approximates the underlying tensor by the sum of a small number of rank-1 tensors, whereas the low Tucker rank model assumes the unfolding matrix along each mode are low rank. To estimate an unknown low-rank tensor from corrupted observations, it is a natural option to consider the rank minimization problem which chooses the tensor of lowest rank as the solution from a certain feasible set. However, tensor rank minimization, even in its 2-way (matrix) case, is generally NP-hard [12] and even harder in higher-way cases [13]. For tractable solutions, researchers turn to a variety of convex surrogates for tensor rank [1,14–18] to replace the tensor rank in rank minimization problem. Methods based on surrogates for the tensor CP rank and Tucker rank have been extensively explored in both the theoretical side and the application side [14,17,19–24].

Recently, the low-tubal-rank model [16,25] has shown better performance than traditional tensor low-rank models in many tensor recover tasks such as image/video inpainting/denoising/sensing [2,25,26], moving object detection [27], multi-view learning [28], seismic data completion [29], WiFi fingerprint [30], MRI imaging [16], point cloud data inpainting [31], and so on. The tubal rank is a new complexity measure of tensor defined through the framework of tensor singular value decomposition (t-SVD) [32,33]. At the core of existing low-tubal-rank models is the tubal nuclear norm (TNN) which is a convex surrogate for the tubal rank. In contrast to CP-based tensor nuclear norms or Tucker-based tensor nuclear norms which models low-rankness in the original domain, TNN models low-rankness in the Fourier domain. It is pointed out in [25,34,35] that TNN has superiority over traditional tensor nuclear norms in exploiting the ubiquitous “spatial-shifting” property in real-world tensor data.

Inspired by the superior performance of TNN, this paper adopts TNN as a low-rank regularizer in the proposed STPCP model. Specifically, the proposed STPCP aims to estimate the underlying tensor data $\underline{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ from an observation tensor $\underline{M}$ polluted by both small dense noises and sparse gross corruptions as follows

$$\underline{M} = \underline{L}_0 + \underline{S}_0 + \underline{E}_0, \quad (1)$$

where $\underline{S}_0$ is a tensor denoting the sparse corruptions and $\underline{E}_0$ is a tensor representing small dense noises. Model (1) is also known as robust tensor decomposition in [36,37].

Our STPCP model is first formulated as a TNN-based convex problem. Then, our theoretical analysis gives upper bound on the estimation error of $\underline{L}_0$ and $\underline{S}_0$. In contrast to the analysis in [37], the proposed STPCP can exactly recovery the underlying tensor $\underline{L}_0$ and the sparse corruption tensor $\underline{S}_0$ when the noise term $\underline{E}_0$ vanishes. For efficient solution of the proposed STPCP model, we develop two algorithms with extensions to a more challenging scenario where missing observations are also considered. The first algorithm is an ADMM algorithm and the second algorithm accelerates it using tensor factorization. Experiments show the effectiveness and the efficiency of the designed algorithms.

We organize the rest of this paper as follows. In Section 2, we briefly introduce basic preliminaries for t-SVD and some related works. The proposed STPCP model is formulated and analyzed theoretically in Section 3. We design two algorithms in Section 4 and report experimental results in Section 5. This work is concluded in Section 6. The proofs of theorems, propositions, and lemmas are given in the appendix.

2. Preliminaries and Related Works

In this section, some preliminaries of t-SVD are first introduced. Then, the related works are presented.

Notations. We denote vectors by bold lower-case letters, e.g., $\mathbf{a} \in \mathbb{R}^n$, matrices by bold upper-case letters, e.g., $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$, and tensors by underlined upper-case letters, e.g., $\underline{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. For a given 3-way tensor, we define its fiber as a vector given through fixing all indices but one, and its slice as a matrix obtained by fixing all indices but two. For a given 3-way tensor $\underline{A}$, we use $\underline{A}_{ijk}$ to

denote its (i, j, k) -th element; $\mathbf{A}^{(k)} := \underline{\mathbf{A}}(:, :, k)$ is used to denote its k -th frontal slice. $\tilde{\underline{\mathbf{A}}}$ is used to denote the tensor after performing 1D Discrete Fourier Transformation (DFT) on all tube fibers $\underline{\mathbf{A}}(i, j, :)$ of $\underline{\mathbf{A}}$, $\forall i = 1, 2, \dots, n_1, j = 1, 2, \dots, n_2$, which can be efficiently computed by the Matlab command $\tilde{\underline{\mathbf{A}}} = \text{fft}(\underline{\mathbf{A}}, [], 3)$. We use $\text{dft3}(\cdot)$ and $\text{idft3}(\cdot)$ to represent the 1D DFT and inverse DFT along the tube fibers of 3-way tensors, i.e., $\text{dft3}(\underline{\mathbf{A}}) := \text{fft}(\underline{\mathbf{A}}, [], 3)$, $\text{idft3}(\underline{\mathbf{A}}) := \text{ifft}(\underline{\mathbf{A}}, [], 3)$.

For a given matrix $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$, define the nuclear norm and spectral norm of $\mathbf{M}$ respectively as

$$\|\mathbf{M}\|_* := \sum_{i=1}^p \sigma_i(\mathbf{M}), \quad \text{and} \quad \|\mathbf{M}\|_{\text{sp}} := \max\{\sigma_i(\mathbf{M})\},$$

where $p = \min\{n_1, n_2\}$, and $\sigma_1(\mathbf{M}) \geq \dots \geq \sigma_p(\mathbf{M})$ are the singular values of $\mathbf{M}$ in a non-ascending order. The l_0 -norm, l_1 -norm, Frobenius norm, l_∞ -norm of a tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as

$$\|\underline{\mathbf{A}}\|_0 := \sum_{ijk} 1(\underline{\mathbf{A}}_{ijk} \neq 0), \quad \|\underline{\mathbf{A}}\|_1 := \sum_{ijk} |\underline{\mathbf{A}}_{ijk}|, \quad \|\underline{\mathbf{A}}\|_F := \sqrt{\sum_{ijk} \underline{\mathbf{A}}_{ijk}^2}, \quad \|\underline{\mathbf{A}}\|_\infty := \max_{ijk} |\underline{\mathbf{A}}_{ijk}|,$$

where $1(C)$ is an indicator function whose value is 1 if the condition C is true, and 0 otherwise.

Given two matrices $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n_1 \times n_2}$, $\mathbf{B} = (b_{ij}) \in \mathbb{C}^{n_1 \times n_2}$, we define their inner product as follows

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^H \mathbf{B}) = \sum_{ij} \bar{a}_{ij} b_{ij},$$

where $\mathbf{A}^H$ denotes conjugate transpose of matrix $\mathbf{A}$ and $\bar{a}_{ij}$ denotes the conjugation of complex number a_{ij} . Given two 3-way tensors $\underline{\mathbf{A}}, \underline{\mathbf{B}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we define their inner product as follows

$$\langle \underline{\mathbf{A}}, \underline{\mathbf{B}} \rangle := \sum_{ijk} \underline{\mathbf{A}}_{ijk} \underline{\mathbf{B}}_{ijk}.$$

2.1. Tensor Singular Value Decomposition

We first define 3 operators based on block matrices which are introduced in [33]. For a given 3-way tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we define its block vectorization $\text{bvec}(\cdot)$ and the inverse operation $\text{bvfold}(\cdot)$ in the following equation:

$$\text{bvec}(\underline{\mathbf{A}}) := \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(n_3)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2}, \quad \text{bvfold}(\text{bvec}(\underline{\mathbf{A}})) = \underline{\mathbf{A}}.$$

We further define the block circulant matrix $\text{bcirc}(\cdot)$ of any 3-way tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as follows:

$$\text{bcirc}(\underline{\mathbf{A}}) := \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(n_3)} & \dots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \dots & \mathbf{A}^{(3)} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{A}^{(n_3)} & \mathbf{A}^{(n_3-1)} & \dots & \mathbf{A}^{(1)} \end{bmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$$

Equipped with above defined operators, we are now in a position to define the t-product of 3-way tensors.

Definition 1 (t-product [33]). Given two tensors $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the t-product of $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ is a new 3-way tensor $\underline{\mathbf{C}}$ with size $n_1 \times n_4 \times n_3$:

$$\underline{\mathbf{C}} = \underline{\mathbf{A}} * \underline{\mathbf{B}} =: \text{bvfold}(\text{bcirc}(\underline{\mathbf{A}}) \text{bvec}(\underline{\mathbf{B}})). \quad (2)$$

A more intuitive interpretation of t-SVD is as follows [33]. If we treat a 3-way tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as a matrix of size $n_1 \times n_2$ whose entries are the tube fibers, then the tensor t-product can be analogously understood as the “matrix multiplication” where the standard scalar product is replaced with the vector circular convolution between the tubes (i.e., vectors):

$$\underline{\mathbf{C}} = \underline{\mathbf{A}} * \underline{\mathbf{B}} \Leftrightarrow \underline{\mathbf{C}}(i, j, :) = \sum_{k=1}^{n_2} \underline{\mathbf{A}}(i, k, :) \star \underline{\mathbf{B}}(k, j, :), \quad \forall i = 1, 2, \dots, n_1, j = 1, 2, \dots, n_3, \quad (3)$$

where $\star$ represent the operation of circular convolution [33] of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n_3}$ defined as $(\mathbf{a} \star \mathbf{b})_j = \sum_{k=1}^{n_3} \mathbf{a}_k \mathbf{b}_{1+(j-k) \bmod n_3}$.

We also define the block diagonal matrix $\text{bdiag}(\cdot)$ of any 3-way tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and its inverse $\text{bdfold}(\cdot)$ as follows

$$\text{bdiag}(\underline{\mathbf{A}}) := \begin{bmatrix} \mathbf{A}^{(1)} & & \\ & \ddots & \\ & & \mathbf{A}^{(n_3)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}, \quad \text{bdfold}(\text{bdiag}(\underline{\mathbf{A}})) = \underline{\mathbf{A}}.$$

We also use $\overline{\mathbf{A}}$ (or $\widetilde{\mathbf{A}}$) to denote the block diagonal matrix of tensor $\widetilde{\mathbf{A}} = \text{dft3}(\underline{\mathbf{A}})$ (i.e., the Fourier version of $\underline{\mathbf{A}}$) i.e.,

$$\overline{\mathbf{A}} = \text{bdiag}(\widetilde{\mathbf{A}}) := \begin{bmatrix} \widetilde{\mathbf{A}}^{(1)} & & \\ & \ddots & \\ & & \widetilde{\mathbf{A}}^{(n_3)} \end{bmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}.$$

Then the relationship between DFT and circular convolution further indicates that the conducting t-product in the original domain is equivalent to performing standard matrix product on the Fourier block diagonal matrices [33]. Since matrix product on the Fourier block diagonal matrices can be parallel written as matrix product of all the frontal slices in the Fourier domain, we have the following relationships

$$\underline{\mathbf{C}} = \underline{\mathbf{A}} * \underline{\mathbf{B}} \Leftrightarrow \overline{\mathbf{C}} = \overline{\mathbf{A}} \overline{\mathbf{B}} \Leftrightarrow \widetilde{\mathbf{C}}^{(k)} = \widetilde{\mathbf{A}}^{(k)} \widetilde{\mathbf{B}}^{(k)}, \quad k = 1, 2, \dots, n_3. \quad (4)$$

The relationship between the t-product and FFT also indicates that the inner product of two 3-way tensors $\underline{\mathbf{A}}, \underline{\mathbf{B}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and the inner product of their corresponding Fourier block diagonal matrices $\overline{\mathbf{A}}, \overline{\mathbf{B}} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$ satisfy the following relationship

$$\langle \underline{\mathbf{A}}, \underline{\mathbf{B}} \rangle = \frac{1}{n_3} \langle \widetilde{\mathbf{A}}, \widetilde{\mathbf{B}} \rangle = \frac{1}{n_3} \langle \overline{\mathbf{A}}, \overline{\mathbf{B}} \rangle.$$

When $\underline{\mathbf{A}} = \underline{\mathbf{B}} = \underline{\mathbf{X}}$, one has

$$\|\underline{\mathbf{X}}\|_F = \frac{1}{\sqrt{n_3}} \|\overline{\mathbf{X}}\|_F.$$

We further define the concepts of tensor transpose, identity tensor, f-diagonal tensor and orthogonal tensor as follows.

Definition 2 (tensor transpose [33]). Given a tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then define its transpose tensor $\underline{\mathbf{A}}^\top$ of size $n_2 \times n_1 \times n_3$ which can be formed through first transposing all the frontal slices of $\underline{\mathbf{A}}$ and then exchanging each k -th transposed frontal slice with the $(n_3 + 2 - k)$ -th transposed frontal slice for all $k = 2, 3, \dots, n_3$.

For example, consider 3-way tensor $\underline{\mathbf{A}} = [\mathbf{A}^{(1)} | \mathbf{A}^{(2)} | \mathbf{A}^{(3)} | \mathbf{A}^{(4)}] \in \mathbb{R}^{n_1 \times n_2 \times 4}$ with 4 frontal slices, the tensor transpose $\underline{\mathbf{A}}^\top$ of $\underline{\mathbf{A}}$ is

$$\underline{\mathbf{A}}^\top = [(\mathbf{A}^{(1)})^\top | (\mathbf{A}^{(4)})^\top | (\mathbf{A}^{(3)})^\top | (\mathbf{A}^{(2)})^\top] \in \mathbb{R}^{n_2 \times n_1 \times 4}.$$

Definition 3 (identity tensor [33]). The identity tensor $\underline{\mathbf{I}} \in \mathbb{R}^{n \times n \times n_3}$ is a tensor whose first frontal slice is the n -by- n identity matrix with all other frontal slices are zero matrices.

Definition 4 (f-diagonal tensor [33]). We call a 3-way tensor f -diagonal if all the frontal slices of it are diagonal matrices.

Definition 5 (orthogonal tensor [33]). We call a tensor $\underline{\mathbf{Q}} \in \mathbb{R}^{n \times n \times n_3}$ an orthogonal tensor if the following equations hold:

$$\underline{\mathbf{Q}}^\top * \underline{\mathbf{Q}} = \underline{\mathbf{Q}} * \underline{\mathbf{Q}}^\top = \underline{\mathbf{I}}.$$

Then, the tensor singular value decomposition (t-SVD) can be given as follows.

Definition 6 (Tensor singular value decomposition, and Tensor tubal rank [38]). Given any 3-way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then it has the following factorization called tensor singular value decomposition (t-SVD)

$$\underline{\mathbf{X}} = \underline{\mathbf{U}} * \underline{\mathbf{\Sigma}} * \underline{\mathbf{V}}^\top, \quad (5)$$

where the left and right factor tensors $\underline{\mathbf{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\underline{\mathbf{V}} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and the middle tensor $\underline{\mathbf{\Sigma}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rectangular f -diagonal tensor.

A visual illustration for the t-SVD is shown in Figure 1. It can be computed efficiently by FFT and IFFT in the Fourier domain according to Equation (4). For more details, see [2].

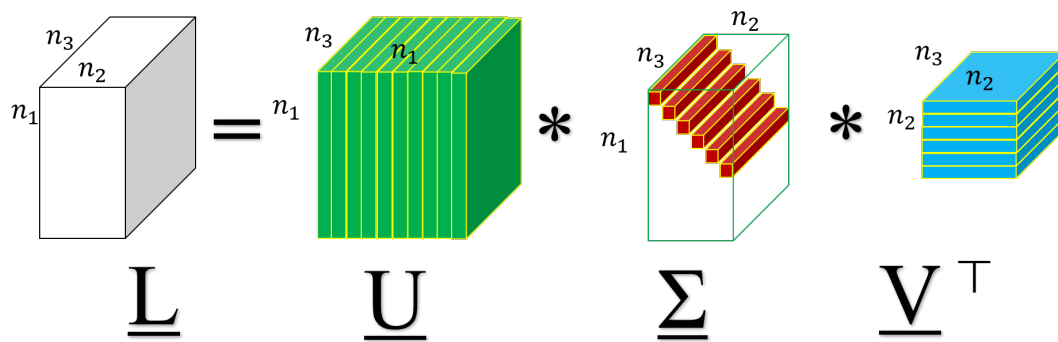


Figure 1. A visual illustration of t-SVD.

Definition 7 (Tensor tubal rank [38]). The tensor tubal rank of any 3-way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the number of non-zero tubes of $\underline{\mathbf{\Sigma}}$ in its t-SVD shown in Equation (5), i.e.,

$$r_{\text{tubal}}(\underline{\mathbf{A}}) := \sum_i 1(\underline{\mathbf{\Sigma}}(i, i, :) \neq \mathbf{0}). \quad (6)$$

Definition 8 (Tubal average rank [38]). The tubal average rank $r_a(\underline{\mathbf{A}})$ of any 3-way tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the averaged rank of all frontal slices of $\underline{\mathbf{A}}$ as follows,

$$r_a(\underline{\mathbf{A}}) := \frac{1}{n_3} \sum_{k=1}^{n_3} \text{rank}(\tilde{\mathbf{A}}^{(k)}). \quad (7)$$

Definition 9 (Tensor operator norm [2,38]). The tensor operator norm $\|\underline{\mathbf{F}}\|_{\text{op}}$ of any 3-way tensor $\underline{\mathbf{F}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as follows

$$\|\underline{\mathbf{F}}\|_{\text{op}} := \sup_{\|\underline{\mathbf{A}}\|_{\text{F}} \leq 1} \|\underline{\mathbf{F}} * \underline{\mathbf{A}}\|_{\text{F}}. \quad (8)$$

The relationship between t-product and FFT indicates that

$$\|\underline{\mathbf{F}}\|_{\text{op}} := \sup_{\|\underline{\mathbf{A}}\|_{\text{F}} \leq 1} \|\underline{\mathbf{F}} * \underline{\mathbf{A}}\|_{\text{F}} = \sup_{\|\underline{\mathbf{A}}\|_{\text{F}} \leq \sqrt{n_3}} \|\underline{\mathbf{F}} \cdot \underline{\mathbf{A}}\|_{\text{F}} = \|\underline{\mathbf{A}}\|_{\text{sp}}. \quad (9)$$

Definition 10 (Tensor spectral norm [38]). The tensor spectral norm $\|\underline{\mathbf{A}}\|_{\text{sp}}$ of any 3-way tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the matrix spectral norm of $\underline{\mathbf{A}}$, i.e.,

$$\|\underline{\mathbf{A}}\|_{\text{sp}} := \|\underline{\mathbf{A}}\|_{\text{sp}}. \quad (10)$$

We further define the tubal nuclear norm.

Definition 11 (Tubal nuclear norm [2]). For any tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with t-SVD $\underline{\mathbf{A}} = \underline{\mathbf{U}} * \underline{\Sigma} * \underline{\mathbf{V}}^{\text{T}}$, the tubal nuclear norm (TNN) of $\underline{\mathbf{A}}$ is defined as

$$\|\underline{\mathbf{A}}\|_{\text{TNN}} := \langle \underline{\Sigma}, \mathbf{I} \rangle = \sum_{i=1}^r \underline{\Sigma}(i, i, 1), \quad (11)$$

where $r = r_{\text{tubal}}(\underline{\mathbf{A}})$.

To understand the tubal nuclear norm, first note that

$$r_{\text{tubal}}(\underline{\mathbf{A}}) = \sum_i 1(\underline{\Sigma}(i, i, :) \neq \mathbf{0}) \stackrel{(i)}{=} \sum_i 1(\underline{\tilde{\Sigma}}(i, i, :) \neq \mathbf{0}) \stackrel{(ii)}{=} \sum_i 1(\|\underline{\tilde{\Sigma}}(i, i, :)\|_1 \neq 0) \stackrel{(iii)}{=} \sum_i 1(\underline{\Sigma}(i, i, 1) \neq 0), \quad (12)$$

where (i) holds because of the definition of DFT [2], (ii) holds by the property of l_1 -norm, and (iii) is a result of DFT [2]. Thus, the tubal rank of $\underline{\mathbf{A}}$ is also the number of non-zero diagonal elements of $\underline{\Sigma}(i, i, 1)$, i.e., the first frontal slice of tensor $\underline{\Sigma}$ in the t-SVD of $\underline{\mathbf{A}}$. Similar to the matrix singular values, the values $\underline{\Sigma}(i, i, 1), i = 1, 2, \dots, n_3$ are also called the singular values of tensor $\underline{\mathbf{A}}$. As the matrix nuclear norm is the sum of matrix singular values, the tubal nuclear norm can be similarly understood as the sum of tensor singular values.

One can also verify by the property of DFT [2] that

$$\|\underline{\mathbf{A}}\|_{\text{TNN}} = \sum_{i=1}^r \underline{\Sigma}(i, i, 1) = \sum_{k=1}^{n_3} \sum_{i=1}^r \underline{\tilde{\Sigma}}(i, i, k) = \frac{1}{n_3} \sum_{k=1}^{n_3} \|\tilde{\mathbf{A}}^{(k)}\|_* = \frac{1}{n_3} \|\underline{\mathbf{A}}\|_*, \quad (13)$$

which indicates that the TNN of $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is also the averaged nuclear norm all frontal slices of $\tilde{\underline{\mathbf{A}}}$. Thus, TNN indeed models the low-rankness of Fourier domain.

Now, we will show that the low-tubal-rank model is ideal to some real-world tensor data, such as color images and videos.

First, we consider a natural image of size $256 \times 256 \times 3$, shown in Figure 2a. In Figure 2b, we plot the distribution of its singular values, i.e., the values of $\underline{\Sigma}(i, i, 1)$ along with the index i . As can be seen from Figure 2b, there are only a small number of singular values with large magnitude, and most of the singular values are close to 0. Then, we can say that some natural color images are approximately low tubal rank.

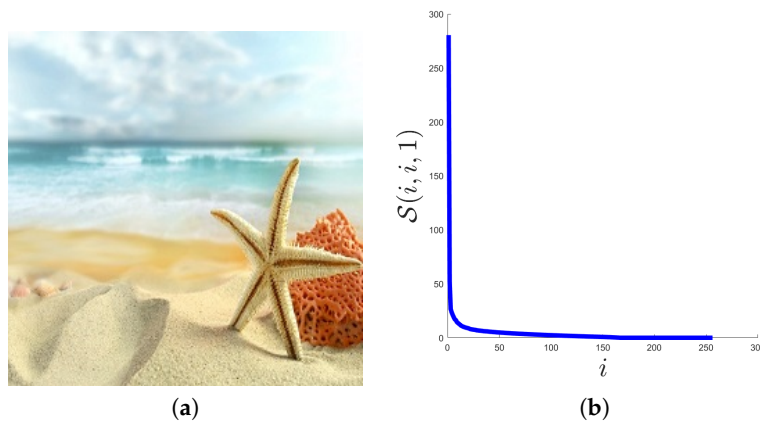


Figure 2. The distribution of tensor singular values $\underline{\Sigma}(i, i, 1)$ in a natural color image. (a) the sample image, (b) the distribution of $\underline{\Sigma}(i, i, 1)$.

Then, consider a commonly used YUV sequence *Mother-daughter_qcif* (These data can be download from the following link <https://sites.google.com/site/subudhibadri/fewhelpfuldownloads>.) whose first frame is shown in Figure 3a. We use the Y components of the first 30 frames, and get a tensor of size $144 \times 176 \times 30$ and show the distribution of tensor singular values in Figure 3b. We can see from Figure 3b that similar to Figure 2b, there are only a small number of singular values with large magnitude, and most of the singular values are close to 0. Then, we can say that some videos can be well approximately low tubal rank.

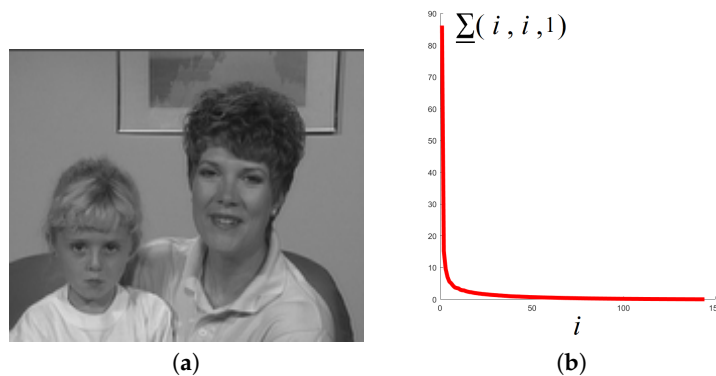


Figure 3. The distribution of tensor singular values $\underline{\Sigma}(i, i, 1)$ in a video sequence. (a) the first frame of the video, (b) the distribution of $\underline{\Sigma}(i, i, 1)$.

For TNN and tensor spectral norm, we highlight the following two lemmas.

Lemma 1. [2] *TNN is the convex envelop of the tensor average rank in the unit ball of tensor spectral norm $\{\underline{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \mid \|\underline{T}\|_{\text{sp}} \leq 1\}$.*

Lemma 2. [2] *The TNN and the tensor spectral norm are dual norms to each other.*

2.2. Related Works

In this subsection, we briefly introduce some related works. The proposed STPCP is tightly related to the Tensor Robust Principal Component Analysis (TRPCA) which aims to recover a low-rank tensor $\underline{L}_0$ and a sparse tensor $\underline{S}_0$ from their sum $\underline{M} = \underline{L}_0 + \underline{S}_0$. This is a special case of our measurement Model (1) where the noise tensor $\underline{E}_0$ is a zero tensor.

In [39], the SNN-based TRPCA model is proposed by modeling the underlying tensor as a low Tucker rank one

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \|\underline{\mathbf{L}}\|_{\text{SNN}} + \|\underline{\mathbf{S}}\|_1 \quad \text{s.t.} \quad \underline{\mathbf{L}} + \underline{\mathbf{S}} = \underline{\mathbf{M}}, \quad (14)$$

where SNN (Sum of Nuclear Norms) is defined as $\|\underline{\mathbf{L}}\|_{\text{SNN}} := \sum_{i=1}^K \alpha_k \|\mathbf{L}_{(k)}\|_*$, where $\alpha_k > 0$ and $\mathbf{L}_{(k)}$ is the mode- k matricization of $\underline{\mathbf{L}}$ [40].

Model (14) indeed assumes the underlying tensor to be low Tucker rank, which can be too strong for some real tensor data. The TNN-based TRPCA model uses TNN to impose low-rankness in the final solution $\underline{\mathbf{L}}$ as follows

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \|\underline{\mathbf{L}}\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}\|_1 \quad \text{s.t.} \quad \underline{\mathbf{L}} + \underline{\mathbf{S}} = \underline{\mathbf{M}}. \quad (15)$$

As shown in [2], when the underlying tensor $\underline{\mathbf{L}}_0$ satisfy the tensor incoherent conditions, by solving Problem (15), one can exactly recover the underlying tensor $\underline{\mathbf{L}}_0$ and $\underline{\mathbf{S}}_0$ with high probability with parameter $\lambda = 1/\sqrt{\max\{n_1, n_2\}n_3}$.

When the noise tensor $\underline{\mathbf{E}}_0$ is not zero, the robust tensor decomposition based on SNN is proposed in [36] as follows

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \frac{1}{2} \|\underline{\mathbf{M}} - \underline{\mathbf{L}} - \underline{\mathbf{S}}\|_F + \lambda_1 \|\underline{\mathbf{L}}\|_{\text{SNN}} + \lambda_2 \|\underline{\mathbf{S}}\|_1, \quad (16)$$

where λ_1 and λ_2 are positive regularization parameters. The estimation error on $\underline{\mathbf{L}}$ and $\underline{\mathbf{S}}$ is analyzed with an upper bound in [36].

In [37], the TNN-based RTD model is proposed as follows

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \frac{1}{2} \|\underline{\mathbf{M}} - \underline{\mathbf{L}} - \underline{\mathbf{S}}\|_F + \lambda_1 \|\underline{\mathbf{L}}\|_{\text{TNN}} + \lambda_2 \|\underline{\mathbf{S}}\|_1, \quad \text{s.t.} \quad \|\underline{\mathbf{L}}\|_\infty \leq \alpha, \quad (17)$$

where α is an upper estimate of l_∞ -norm of the underlying tensor $\underline{\mathbf{L}}_0$. An upper bound on the estimation error is also established. However, in the analysis of Model (17), the error does not vanish as the noise tensor $\underline{\mathbf{E}}_0$ vanishes which means the analysis cannot guarantee exact recovery in the noiseless setting (which can be provided by the analysis of TNN-based TRPCA (15) by Lu et al. [2]).

The Bayesian approach is also used for robust tensor recovery. The CP decomposition under sparse corruption and small dense noise is considered [41], and tensor rank estimation is achieved using Bayesian approach. In [42], CP decomposition under missing value and small dense noise is considered with rank estimation similar to [41]. A sparse Bayesian CP model is proposed in [43] to recover a tensor with missing value, outliers and noises. In [44], a fully Bayesian treatment is proposed to recover a low-tubal-rank tensor corrupted by both noises and outliers.

3. Theoretical Guarantee for Stable Tensor Principal Component Pursuit

In this section, we formulate the proposed STPCP model and give the main theoretical result which upper bounds the estimation error and guarantees exact recovery in the noiseless setting.

3.1. The Proposed STPCP

As for the measurement Model (1), we further assume that the noise tensor $\underline{\mathbf{E}}_0$ has bounded energy measured in F-norm, i.e., $\|\underline{\mathbf{E}}_0\|_F \leq \delta$. Please note that the limited energy assumption is very mild, since most signals are of limited energy.

To recover the low-rank tensor $\underline{\mathbf{L}}_0$ and the sparse tensor $\underline{\mathbf{S}}_0$, we first produce the following optimization problem:

$$(\hat{\underline{\mathbf{L}}}, \hat{\underline{\mathbf{S}}}) = \underset{\underline{\mathbf{L}}, \underline{\mathbf{S}}}{\operatorname{argmin}} \|\underline{\mathbf{L}}\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}\|_1, \quad \text{s.t.} \quad \|\underline{\mathbf{M}} - \underline{\mathbf{L}} - \underline{\mathbf{S}}\|_F \leq \delta, \quad (18)$$

where λ is a positive parameter balancing the two regularizers. The motivation is to use TNN as a low-rank regularization term to exploit the low-dimensional structure in the signal tensor, whereas tensor l_1 -norm is used to impose sparsity in the corruption tensor (since we assume it to be sparse).

The relationship between Model (18) and existing models are discussed in Remark 1 and Remark 2.

Remark 1. The following models can be seen as special cases as the proposed STPCP Model (18);

- (I). When $\delta = 0$, i.e., in the noiseless case, the proposed model degenerates to the TRPCA Model (15) [2].
- (II). When $n_3 = 1$, then the stable tensor PCP Model (18) degenerates to the Stable Principal Component Pursuit (SPCP) [45] which aims to pursue the principal components modeled by low-rank matrix $\underline{L}_0$ from its observation $\mathbf{M}$ corrupted by both noises $\mathbf{E}_0$ and sparse corruptions $\mathbf{S}_0$. The SPCP is formulated as follows

$$\min_{\underline{L}, \mathbf{S}} \|\underline{L}\|_* + \lambda \|\mathbf{S}\|_1, \quad \text{s.t. } \|\mathbf{M} - \underline{L} - \mathbf{S}\|_F \leq \delta. \quad (19)$$

- (III). When $n_3 = 1$ and $\delta = 0$, the proposed STPCP further degenerates to Robust Principal Component Analysis (RPCA) [46] given as follows

$$\min_{\underline{L}, \mathbf{S}} \|\underline{L}\|_* + \lambda \|\mathbf{S}\|_1, \quad \text{s.t. } \underline{L} + \mathbf{S} = \mathbf{M}. \quad (20)$$

Remark 2. The differences from the proposed Model (18) and TNN-based RTD Model (17 [37]) is as follows. First, our model does not need to upper estimate the l_∞ -norm of the underlying tensor. Second, our model is a constrained optimization problem, whereas Model (17) is an unconstrained optimization problem.

3.2. A Theorem for Stable Recovery

To analyze the statistical performance of Model (18), we should assume on the underlying low-rank tensor $\underline{L}_0$ that it is not sparse. Only by this assumption, $\underline{L}_0$ can be identified from its mixture with sparse $\underline{S}_0$. Such an assumption can be described by the tensor incoherence condition [2,47], which is used to provide an identifiability for low-rank $\underline{L}_0$.

Definition 12 (Tensor incoherence condition [2,47]). Given a 3-way tensor $\underline{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r , suppose it has the skinny t -SVD $\underline{T} = \underline{U} * \underline{\Lambda} * \underline{V}^\top$, where $\underline{U} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\underline{V} \in \mathbb{R}^{r \times n_2 \times n_3}$ are orthogonal tensors, and $\underline{\Lambda} \in \mathbb{R}^{r \times r \times n_3}$ is an f -diagonal tensor. Then, $\underline{T}$ is said to satisfy the tensor incoherent condition (TIC) with parameter $\mu(\underline{T})$ if the following inequalities hold

$$\max_{i \in [n_1]} \|\underline{U}^\top * \mathbf{\hat{e}}_i\|_F \leq \sqrt{\frac{r\mu(\underline{T})}{n_1 n_3}}, \quad (21)$$

$$\max_{j \in [n_2]} \|\underline{V}^\top * \mathbf{\hat{e}}_j\|_F \leq \sqrt{\frac{r\mu(\underline{T})}{n_2 n_3}}, \quad (22)$$

$$\|\underline{U} * \underline{V}^\top\|_\infty \leq \sqrt{\frac{r\mu(\underline{T})}{n_1 n_2 n_3}}. \quad (23)$$

where $\mathbf{\hat{e}}_i \in \mathbb{R}^{n_1 \times 1 \times n_3}$ is a tensor column basis with only the $(i, 1, 1)$ -th element being 1 and all the others being 0, and $\mathbf{\hat{e}}_j \in \mathbb{R}^{n_2 \times 1 \times n_3}$ is also a tensor column basis with only the $(j, 1, 1)$ -th element being 1 and all the others being 0.

Assumption 1. Suppose the true tensor $\underline{L}_0$ in the measurement model (1) satisfies tensor incoherence condition with parameter μ .

Assumption 1 intrinsically ensures that the row bases and column bases of $\underline{L}_0$ do not align well with the canonical row and column bases. Thus, the low-rank $\underline{L}_0$ is not sparse, which avoids the ambiguity that low-rank component can also be sparse in the measurement Model (1).

We should also force the sparse component in Model (1) is not low rank.

Assumption 2. Assume the support Ω of $\underline{S}_0$ is drawn uniformly at random.

Now we can establish an upper bound on the estimation error of $\underline{\hat{L}}$ and $\underline{\hat{S}}$ in Problem (18).

Theorem 1 (An Upper Bound on the Estimation Error). Suppose $\underline{L}_0$ and $\underline{S}_0$ satisfy Assumption 1 and Assumption 2, respectively. If the tubal rank r of $\underline{L}_0$ and the sparsity (i.e., the l_0 -norm) s of $\underline{S}_0$ are respectively upper bounded as follows

$$r \leq \frac{c_r \min\{n_1, n_2\}}{\mu \log^2(n_3 \max\{n_1, n_2\})}, \quad \text{and} \quad s \leq c_s n_1 n_2 n_3 \quad (24)$$

where c_l and c_s are two sufficiently small numerical constants independent on the dimensions n_1, n_2 and n_3 . Then the estimator defined in Model (18) satisfy the following inequalities

$$\begin{aligned} \|\underline{\hat{L}} - \underline{L}_0\|_F &\leq \left(\sqrt{1 + \frac{1}{\max\{n_1, n_2\}}} + 8(1 + 2\sqrt{2})\sqrt{\min\{n_1, n_2\}n_3} \right) \delta \\ \|\underline{\hat{S}} - \underline{S}_0\|_F &\leq \left(\sqrt{1 + \max\{n_1, n_2\}} + 8(1 + 2\sqrt{2})\sqrt{n_1 n_2 n_3} \right) \delta, \end{aligned} \quad (25)$$

with probability at least $1 - c_1(n_3 \max\{n_1, n_2\})^{-c_2}$ (over the choice of support of $\underline{S}_0$), where c_1 and c_2 are positive constants independent on the dimensions n_1, n_2 and n_3 .

The proof of Theorem 1 are given in the appendix. In Theorem 1, estimation errors on $\underline{L}_0$ and $\underline{S}_0$ are separately established. It indicates that the estimation error scales linearly with the noise level δ , which is in consistence with the result in [37].

Remark 3. A significant progress over [37] is that in the noiseless setting where $\underline{E}_0$ vanishes, our analysis can provide exact recovery guarantee of $\underline{L}_0$ and $\underline{S}_0$. This is because the tensor incoherence condition adopted in our analysis intrinsically ensures that the low-rank tensor $\underline{L}_0$ is not sparse and thus can be separated from the sparse corruption tensor, whereas the non-spiky condition adopted in [37] fails to provide identifiability in the measurement Model (1).

For Theorem 1, we also give the following remark.

Remark 4. The error bounds established in Theorem 1 are consistent with the theoretical analysis for the special cases shown in Remark 1.

- (I). When $\delta = 0$, i.e., in the noiseless case, the error bounds in Theorem 1 will vanish, which means exact recovery of $\underline{L}_0$ and $\underline{S}_0$ can be guaranteed. This result is consistent with the analysis in [2] for TNN-based TRPCA Model (15).
- (II). When $n_3 = 1$, the error bound on the sparse component in Theorem 1 is consistent with the error bound shown in Equation (8) of [45]. The upper bound on error of the low-rank component in Theorem 1 is sharper than that given in Equation (8) of [45].
- (III). When $n_3 = 1$ and $\delta = 0$, the proposed STPCP has consistent theoretical guarantee with the analysis of RPCA [46].

4. Algorithms

In this section, we design two algorithms. The first algorithm is based on the framework of ADMM [48] which has been extensively used in convex optimization with good convergence behavior. However, ADMM requires full SVDs on large matrices in each iteration which is high computational burden in high-dimensional settings. Thus, the second algorithm is proposed to solve this issue by using a factorization trick which can instead conducting SVDs on much smaller matrices.

4.1. An ADMM Algorithm

The proposed estimator (18) is equivalent to the following unconstrained problem:

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \frac{1}{2} \|\underline{\mathbf{L}} + \underline{\mathbf{S}} - \underline{\mathbf{M}}\|_{\mathbf{F}}^2 + \gamma(\|\underline{\mathbf{L}}\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}\|_1), \quad (26)$$

where γ is a positive parameter balancing the data fidelity term and the regularization term.

Besides being corrupted by noises and outliers, the observed tensor $\underline{\mathbf{M}}$ may also suffer from missing entries which can be taken as outliers with known positions in many applications. Thus, it is more practical to consider the recovery of $\underline{\mathbf{L}}_0$ against outliers $\underline{\mathbf{S}}_0$, noises $\underline{\mathbf{E}}_0$ and missing entries shown in the following measurement model

$$\underline{\mathbf{M}} = \underline{\mathbf{B}} \odot (\underline{\mathbf{L}}_0 + \underline{\mathbf{S}}_0 + \underline{\mathbf{E}}_0), \quad (27)$$

where tensor $\underline{\mathbf{B}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denote the missing mask where $\underline{\mathbf{B}}_{ijk} = 1$, if the (i, j, k) -th entry of $\underline{\mathbf{L}}$ is observed and $\underline{\mathbf{B}}_{ijk} = 0$ otherwise, and $\odot$ denotes element-wise multiplication. Taking into consideration of missing entries, Model (26) can be further modified as:

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{L}} + \underline{\mathbf{S}} - \underline{\mathbf{M}})\|_{\mathbf{F}}^2 + \gamma(\|\underline{\mathbf{L}}\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}\|_1). \quad (28)$$

By adding auxiliary variables to Problem (28), we obtain

$$\begin{aligned} \min_{\underline{\mathbf{K}}, \underline{\mathbf{L}}, \underline{\mathbf{R}}, \underline{\mathbf{S}}} \quad & \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{L}} + \underline{\mathbf{S}} - \underline{\mathbf{M}})\|_{\mathbf{F}}^2 + \gamma \|\underline{\mathbf{K}}\|_{\text{TNN}} + \gamma \lambda \|\underline{\mathbf{R}}\|_1 \\ \text{s.t.} \quad & \underline{\mathbf{K}} = \underline{\mathbf{L}}, \quad \underline{\mathbf{R}} = \underline{\mathbf{S}}. \end{aligned} \quad (29)$$

The Augmented Lagrangian (AL) of Problem (29) is given as follows

$$\begin{aligned} L_{\rho}(\underline{\mathbf{L}}, \underline{\mathbf{S}}, \underline{\mathbf{K}}, \underline{\mathbf{R}}, \underline{\mathbf{Y}}_1, \underline{\mathbf{Y}}_2) = & \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{L}} + \underline{\mathbf{S}} - \underline{\mathbf{M}})\|_{\mathbf{F}}^2 + \gamma \|\underline{\mathbf{K}}\|_{\text{TNN}} + \gamma \lambda \|\underline{\mathbf{R}}\|_1 \\ & + \langle \underline{\mathbf{Y}}_1, \underline{\mathbf{K}} - \underline{\mathbf{L}} \rangle + \frac{\rho}{2} \|\underline{\mathbf{K}} - \underline{\mathbf{L}}\|_{\mathbf{F}}^2 + \langle \underline{\mathbf{Y}}_2, \underline{\mathbf{R}} - \underline{\mathbf{S}} \rangle + \frac{\rho}{2} \|\underline{\mathbf{R}} - \underline{\mathbf{S}}\|_{\mathbf{F}}^2, \end{aligned} \quad (30)$$

where $\underline{\mathbf{Y}}_1, \underline{\mathbf{Y}}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are Lagrangian multipliers and ρ is a penalty parameter.

According the strategy of ADMM, we update prime variables $(\underline{\mathbf{L}}, \underline{\mathbf{S}})$ and $(\underline{\mathbf{K}}, \underline{\mathbf{R}})$ by alternative minimization of AL in Problem (29) as follows

- Update $(\underline{\mathbf{L}}, \underline{\mathbf{S}})$. We update $(\underline{\mathbf{L}}, \underline{\mathbf{S}})$ by minimizing L_{ρ} with other variables fixed as follows

$$\begin{aligned} & (\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{S}}^{t+1}) \\ & = \operatorname{argmin}_{(\underline{\mathbf{L}}, \underline{\mathbf{S}})} L_{\rho}(\underline{\mathbf{L}}, \underline{\mathbf{S}}, \underline{\mathbf{K}}^t, \underline{\mathbf{R}}^t, \underline{\mathbf{Y}}_1^t, \underline{\mathbf{Y}}_2^t) \\ & = \operatorname{argmin}_{(\underline{\mathbf{L}}, \underline{\mathbf{S}})} \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{L}} + \underline{\mathbf{S}} - \underline{\mathbf{M}})\|_{\mathbf{F}}^2 + \langle \underline{\mathbf{Y}}_1^t, \underline{\mathbf{K}}^t - \underline{\mathbf{L}} \rangle + \frac{\rho}{2} \|\underline{\mathbf{K}}^t - \underline{\mathbf{L}}\|_{\mathbf{F}}^2 + \langle \underline{\mathbf{Y}}_2^t, \underline{\mathbf{R}}^t - \underline{\mathbf{S}} \rangle + \frac{\rho}{2} \|\underline{\mathbf{R}}^t - \underline{\mathbf{S}}\|_{\mathbf{F}}^2. \end{aligned} \quad (31)$$

Taking derivatives of the right-hand side of Equation (31) with respect to $\underline{\mathbf{L}}$ and $\underline{\mathbf{S}}$ respectively, and setting the results zero, we obtain

$$\begin{aligned} \underline{\mathbf{B}} \odot (\underline{\mathbf{L}}^{t+1} + \underline{\mathbf{S}}^{t+1}) - \underline{\mathbf{B}} \odot \underline{\mathbf{M}} - \underline{\mathbf{Y}}_1^t + \rho(\underline{\mathbf{L}}^{t+1} - \underline{\mathbf{K}}^t) &= 0 \\ \underline{\mathbf{B}} \odot (\underline{\mathbf{L}}^{t+1} + \underline{\mathbf{S}}^{t+1}) - \underline{\mathbf{B}} \odot \underline{\mathbf{M}} - \underline{\mathbf{Y}}_2^t + \rho(\underline{\mathbf{S}}^{t+1} - \underline{\mathbf{R}}^t) &= 0. \end{aligned} \quad (32)$$

Resolving the above equation group yields

$$\begin{aligned}\underline{\mathbf{L}}^{t+1} &= (\rho(\underline{\mathbf{B}} + \rho \underline{\mathbf{1}}) \odot \underline{\mathbf{K}}^t + \rho \underline{\mathbf{B}} \odot \underline{\mathbf{M}} + (\underline{\mathbf{B}} + \rho \underline{\mathbf{1}}) \odot \underline{\mathbf{Y}}_1^t - \underline{\mathbf{B}} \odot \underline{\mathbf{Y}}_2^t - \rho \underline{\mathbf{B}} \odot \underline{\mathbf{R}}^t) \oslash (\rho(2\underline{\mathbf{B}} + \rho \underline{\mathbf{1}})), \\ \underline{\mathbf{S}}^{t+1} &= (\rho(\underline{\mathbf{B}} + \rho \underline{\mathbf{1}}) \odot \underline{\mathbf{R}}^t + \rho \underline{\mathbf{B}} \odot \underline{\mathbf{M}} + (\underline{\mathbf{B}} + \rho \underline{\mathbf{1}}) \odot \underline{\mathbf{Y}}_2^t - \underline{\mathbf{B}} \odot \underline{\mathbf{Y}}_1^t - \rho \underline{\mathbf{B}} \odot \underline{\mathbf{K}}^t) \oslash (\rho(2\underline{\mathbf{B}} + \rho \underline{\mathbf{1}})),\end{aligned}\quad (33)$$

where $\oslash$ denotes entry-wise division and $\underline{\mathbf{1}}$ denotes the tensor all whose entries are 1.

- Update $(\underline{\mathbf{K}}, \underline{\mathbf{R}})$. We update $(\underline{\mathbf{K}}, \underline{\mathbf{R}})$ by minimizing L_ρ with other variables fixed as follows

$$\begin{aligned}(\underline{\mathbf{K}}^{t+1}, \underline{\mathbf{R}}^{t+1}) &= \underset{(\underline{\mathbf{K}}, \underline{\mathbf{R}})}{\operatorname{argmin}} L_\rho(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{S}}^{t+1}, \underline{\mathbf{K}}, \underline{\mathbf{R}}, \underline{\mathbf{Y}}_1^t, \underline{\mathbf{Y}}_2^t) \\ &= \underset{(\underline{\mathbf{K}}, \underline{\mathbf{R}})}{\operatorname{argmin}} \gamma \|\underline{\mathbf{K}}\|_{\text{TNN}} + \gamma \lambda \|\underline{\mathbf{R}}\|_1 + \left\langle \underline{\mathbf{Y}}_1^t, \underline{\mathbf{K}} - \underline{\mathbf{L}}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{\mathbf{K}} - \underline{\mathbf{L}}^{t+1}\|_{\text{F}}^2 + \left\langle \underline{\mathbf{Y}}_2^t, \underline{\mathbf{R}} - \underline{\mathbf{S}}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{\mathbf{R}} - \underline{\mathbf{S}}^{t+1}\|_{\text{F}}^2.\end{aligned}\quad (34)$$

Please note that Problem (34) can further be solved separately as follows

$$\begin{aligned}\underline{\mathbf{K}}^{t+1} &= \underset{\underline{\mathbf{K}}}{\operatorname{argmin}} \gamma \|\underline{\mathbf{K}}\|_{\text{TNN}} + \left\langle \underline{\mathbf{Y}}_1^t, \underline{\mathbf{K}} - \underline{\mathbf{L}}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{\mathbf{K}} - \underline{\mathbf{L}}^{t+1}\|_{\text{F}}^2 \\ &= \mathfrak{S}_{\gamma \rho^{-1}}^{\|\cdot\|_{\text{TNN}}} \left(\underline{\mathbf{L}}^{t+1} - \rho^{-1} \underline{\mathbf{Y}}_1^t \right).\end{aligned}\quad (35)$$

and

$$\begin{aligned}\underline{\mathbf{R}}^{t+1} &= \underset{\underline{\mathbf{R}}}{\operatorname{argmin}} \gamma \lambda \|\underline{\mathbf{R}}\|_1 + \left\langle \underline{\mathbf{Y}}_2^t, \underline{\mathbf{R}} - \underline{\mathbf{S}}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{\mathbf{R}} - \underline{\mathbf{S}}^{t+1}\|_{\text{F}}^2 \\ &= \mathfrak{S}_{\gamma \lambda \rho^{-1}}^{\|\cdot\|_1} \left(\underline{\mathbf{S}}^{t+1} - \rho^{-1} \underline{\mathbf{Y}}_2^t \right),\end{aligned}\quad (36)$$

where $\mathfrak{S}_{\tau}^{\|\cdot\|_{\text{TNN}}}(\cdot)$ is the proximity operator of TNN [5], and $\mathfrak{S}_{\tau}^{\|\cdot\|_1}(\cdot)$ is the proximity operator of tensor l_1 -norm given as follows [49]:

$$\mathfrak{S}_{\tau}^{\|\cdot\|_1}(\underline{\mathbf{A}}) := \underset{\underline{\mathbf{X}}}{\operatorname{argmin}} \tau \|\underline{\mathbf{X}}\|_1 + \frac{1}{2} \|\underline{\mathbf{X}} - \underline{\mathbf{A}}\|_{\text{F}}^2 = \operatorname{sign}(\underline{\mathbf{A}}) \odot \max\{(|\underline{\mathbf{A}}| - \tau, 0),$$

In [5], a closed-form expression of $\mathfrak{S}_{\tau}(\cdot)$ is given as follows:

Lemma 3 (Proximity operator of TNN [5]). *For any 3D tensor $\underline{\mathbf{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with reduced t -SVD $\underline{\mathbf{A}} = \underline{\mathbf{U}} * \underline{\mathbf{\Lambda}} * \underline{\mathbf{V}}^{\top}$, where $\underline{\mathbf{U}} \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\underline{\mathbf{V}} \in \mathbb{R}^{n_2 \times r \times n_3}$ are orthogonal tensors and $\underline{\mathbf{\Lambda}} \in \mathbb{R}^{r \times r \times n_3}$ is the f -diagonal tensor of singular tubes, the proximity operator $\mathfrak{S}_{\tau}^{\|\cdot\|_{\text{TNN}}}(\underline{\mathbf{A}})$ at $\underline{\mathbf{A}}$ can be computed by:*

$$\mathfrak{S}_{\tau}^{\|\cdot\|_{\text{TNN}}}(\underline{\mathbf{A}}) := \underset{\underline{\mathbf{X}}}{\operatorname{argmin}} \tau \|\underline{\mathbf{X}}\|_{\text{TNN}} + \frac{1}{2} \|\underline{\mathbf{X}} - \underline{\mathbf{A}}\|_{\text{F}}^2 = \underline{\mathbf{U}} * \operatorname{fft3}(\max(\operatorname{fft3}(\underline{\mathbf{\Lambda}}) - \tau, 0)) * \underline{\mathbf{V}}^{\top},$$

- Update $(\underline{\mathbf{Y}}_1, \underline{\mathbf{Y}}_2)$. The Lagrangian multipliers are updated by gradient ascent as follows

$$\begin{aligned}\underline{\mathbf{Y}}_1^{t+1} &= \underline{\mathbf{Y}}_1^t + \rho(\underline{\mathbf{K}}^{t+1} - \underline{\mathbf{L}}^{t+1}), \\ \underline{\mathbf{Y}}_2^{t+1} &= \underline{\mathbf{Y}}_2^t + \rho(\underline{\mathbf{R}}^{t+1} - \underline{\mathbf{S}}^{t+1}).\end{aligned}\quad (37)$$

The algorithm is summarized in Algorithm 1. The convergence analysis of Algorithm 1 is established in Theorem 2.

Algorithm 1 Solving Problem (29) using ADMM.**Input:** The observed tensor $\underline{\mathbf{M}}$, the parameters $\gamma, \lambda, \rho, \delta$.1: Initialize $t = 0, \underline{\mathbf{L}}^0 = \underline{\mathbf{S}}^0 = \underline{\mathbf{K}}^0 = \underline{\mathbf{R}}^0 = \underline{\mathbf{Y}}_1^0 = \underline{\mathbf{Y}}_2^0 = \underline{\mathbf{0}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 2: **for** $t = 0, \dots, T_{\max}$ **do**3: Update $(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{S}}^{t+1})$ by Equation (33);4: Update $(\underline{\mathbf{K}}^{t+1}, \underline{\mathbf{R}}^{t+1})$ by Equations (35)–(36);5: Update $(\underline{\mathbf{Y}}_1^{t+1}, \underline{\mathbf{Y}}_2^{t+1})$ by Equation (37);

6: Check the convergence criteria:

(i) convergence of variables: $\|\underline{\mathbf{A}}^{t+1} - \underline{\mathbf{A}}^t\|_\infty \leq \delta, \forall \underline{\mathbf{A}} \in \{\underline{\mathbf{L}}, \underline{\mathbf{S}}, \underline{\mathbf{K}}, \underline{\mathbf{R}}\}$,(ii) convergence of constraints: $\max\{\|\underline{\mathbf{K}}^{t+1} - \underline{\mathbf{L}}^t\|_\infty, \|\underline{\mathbf{R}}^{t+1} - \underline{\mathbf{S}}^{t+1}\|_\infty\} \leq \delta$.7: **end for****Output:** $(\hat{\underline{\mathbf{L}}}, \hat{\underline{\mathbf{S}}}) = (\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{S}}^{t+1})$.

Theorem 2 (Convergence of Algorithm 1). For any $\rho > 0$, if the unaugmented Lagrangian $L(\underline{\mathbf{L}}, \underline{\mathbf{S}}, \underline{\mathbf{K}}, \underline{\mathbf{R}}, \underline{\mathbf{Y}}_1, \underline{\mathbf{Y}}_2)$ has a saddle point, then the iterations $L(\underline{\mathbf{L}}^t, \underline{\mathbf{S}}^t, \underline{\mathbf{K}}^t, \underline{\mathbf{R}}^t, \underline{\mathbf{Y}}_1^t, \underline{\mathbf{Y}}_2^t)$ in Algorithm 1 satisfy the residual convergence, objective convergence and dual variable convergence of Problem (29) as $t \rightarrow \infty$.

The proof of Theorem 2 is given in the Appendix A.

In a single iteration of Algorithm 1, the main cost comes from updating $\underline{\mathbf{L}}^t$ which involves computing FFT, IFFT and n_3 SVDs of $n_1 \times n_2$ matrices [47]. Hence Algorithm 1 has per-iteration complexity of order $O(n_1 n_2 n_3 (n_1 \wedge n_2 + \log n_3))$. Thus, if the total iteration number is T , then the total computational complexity is

$$O(T n_1 n_2 n_3 (\min\{n_1, n_2\} + \log n_3)). \quad (38)$$

4.2. A Faster Algorithm

To reduce the cost of computing TNN which is a main cost of Algorithm 1, we propose the following lemma which indicates that TNN is orthogonal invariant.

Lemma 4. Given a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{r \times r \times n_3}$, let $\underline{\mathbf{Q}} \in \mathbb{R}^{n_1 \times r \times n_3}$ a two semi-orthogonal tensors, i.e., $\underline{\mathbf{Q}}^\top * \underline{\mathbf{Q}} = \underline{\mathbf{I}} \in \mathbb{R}^{r \times r \times n_3}$ and $r \leq \min\{n_1, n_2\}$. Then, we have the following relationship:

$$\|\underline{\mathbf{Q}} * \underline{\mathbf{X}}\|_{\text{TNN}} = \|\underline{\mathbf{X}}\|_{\text{TNN}}.$$

The proof of Lemma 4 can be found in the appendix. Equipped with Lemma 4, we decompose the low-rank component in Problem (28) as follows:

$$\underline{\mathbf{L}} = \underline{\mathbf{Q}} * \underline{\mathbf{X}}, \quad \text{s.t. } \underline{\mathbf{Q}}^\top * \underline{\mathbf{Q}} = \underline{\mathbf{I}}_r,$$

where $\underline{\mathbf{I}}_r \in \mathbb{R}^{r \times r \times n_3}$ is an identity tensor. The similar strategy has been used in low-rank matrix recovery from gross corruptions by [50]. Furthermore, we propose the following model for Problem (28)

$$\begin{aligned} \min_{\underline{\mathbf{Q}}, \underline{\mathbf{X}}, \underline{\mathbf{S}}} & \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{Q}} * \underline{\mathbf{X}} + \underline{\mathbf{S}} - \underline{\mathbf{M}})\|_F^2 + \gamma (\|\underline{\mathbf{X}}\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}\|_1) \\ \text{s.t. } & \underline{\mathbf{Q}}^\top * \underline{\mathbf{Q}} = \underline{\mathbf{I}}_r, \end{aligned} \quad (39)$$

where r is an upper estimation of tubal rank of the underlying tensor $r^* = r_{\text{tubal}}(\underline{\mathbf{L}}_0)$.

In contrast to Model (28), the proposed Model (39) is a non-convex optimization problem. That means Model (39) may have many local minima. We establish a connection between the proposed Model (39) with Model (28) in the following theorem.

Theorem 3 (Connection between Model (39) and Model (28)). Let $(\underline{Q}_*, \underline{X}_*, \underline{S}_*)$ be a global optimal solution to Problem (39). Furthermore, let $(\underline{L}^*, \underline{S}^*)$ be the solution to Problem (28), and $r_{\text{tubal}}(\underline{L}^*) \leq r$, where r is the initialized tubal rank. Then $(\underline{Q}_* * \underline{X}_*, \underline{S}_*)$ is also the optimal solution to Problem (28).

The proof of Theorem 3 can be found in the appendix. Theorem 3 states that the global optimal point of the (non-convex) Model (39) coincides with solution of the (convex) Model (28). It further indicates that the accuracy of Model (39) cannot exceed Model (28), which can be validated numerically in the experiment section.

To solve Model (39), we also use the ADMM framework.

First, by adding auxiliary variables, we have the following problem

$$\begin{aligned} \min_{\underline{L}, \underline{S}, \underline{R}, \underline{Q}, \underline{X}} \quad & \frac{1}{2} \|\underline{B} \odot (\underline{L} + \underline{S} - \underline{M})\|_F^2 + \gamma(\|\underline{X}\|_{\text{TNN}} + \lambda \|\underline{R}\|_1) \\ \text{s.t.} \quad & \underline{Q} * \underline{X} = \underline{L}; \quad \underline{R} = \underline{S}; \quad \underline{Q}^\top * \underline{Q} = \underline{I}_r. \end{aligned} \quad (40)$$

The augmented Lagrangian of Problem (40) is

$$\begin{aligned} L'_2(\underline{L}, \underline{S}, \underline{R}, \underline{Q}, \underline{X}) = & \frac{1}{2} \|\underline{B} \odot (\underline{L} + \underline{S} - \underline{M})\|_F^2 + \gamma(\|\underline{X}\|_{\text{TNN}} + \lambda \|\underline{R}\|_1) \\ & + \langle \underline{Y}_1, \underline{Q} * \underline{X} - \underline{L} \rangle + \frac{\rho}{2} \|\underline{Q} * \underline{X} - \underline{L}\|_F^2 + \langle \underline{Y}_2, \underline{R} - \underline{S} \rangle + \frac{\rho}{2} \|\underline{R} - \underline{S}\|_F^2 \\ \text{s.t.} \quad & \underline{Q}^\top * \underline{Q} = \underline{I}_r. \end{aligned} \quad (41)$$

According the strategy of ADMM, we update prime variables $(\underline{L}, \underline{S})$ and $(\underline{Q}, \underline{X}, \underline{R})$ by alternative minimization of AL in Problem (41) as follows

- Update $(\underline{L}, \underline{S})$: We update $(\underline{L}, \underline{S})$ by minimizing L'_ρ with other variables fixed as follows

$$\begin{aligned} & (\underline{L}^{t+1}, \underline{S}^{t+1}) \\ & = \arg\min_{(\underline{L}, \underline{S})} L'_\rho(\underline{L}, \underline{S}, \underline{Q}^t, \underline{X}^t, \underline{R}^t, \underline{Y}_1^t, \underline{Y}_2^t) \\ & = \arg\min_{(\underline{L}, \underline{S})} \frac{1}{2} \|\underline{B} \odot (\underline{L} + \underline{S} - \underline{M})\|_F^2 + \langle \underline{Y}_1^t, \underline{Q}^t * \underline{X}^t - \underline{L} \rangle + \frac{\rho}{2} \|\underline{Q}^t * \underline{X}^t - \underline{L}\|_F^2 + \langle \underline{Y}_2^t, \underline{R}^t - \underline{S} \rangle + \frac{\rho}{2} \|\underline{R}^t - \underline{S}\|_F^2. \end{aligned} \quad (42)$$

Taking derivatives of the right-hand side with respect to $\underline{L}$ and $\underline{S}$ respectively, and setting the results zero, we obtain

$$\begin{aligned} \underline{B} \odot (\underline{L}^{t+1} + \underline{S}^{t+1}) - \underline{B} \odot \underline{M} - \underline{Y}_1^t + \rho(\underline{L}^{t+1} - \underline{Q}^t * \underline{X}^t) &= \underline{0} \\ \underline{B} \odot (\underline{L}^{t+1} + \underline{S}^{t+1}) - \underline{B} \odot \underline{M} - \underline{Y}_2^t + \rho(\underline{S}^{t+1} - \underline{R}^t) &= \underline{0}, \end{aligned} \quad (43)$$

Resolving the above equation group yields

$$\begin{aligned} \underline{L}^{t+1} &= ((1 + \rho)\underline{Q}^t * \underline{X}^t + \underline{B} \odot \underline{M} + \underline{Y}_1^t - \underline{R}^t) \oslash (2\underline{B} + \rho\underline{1}), \\ \underline{S}^{t+1} &= ((1 + \rho)\underline{R}^t + \underline{B} \odot \underline{M} + \underline{Y}_2^t - \underline{Q}^t * \underline{X}^t) \oslash (2\underline{B} + \rho\underline{1}). \end{aligned} \quad (44)$$

- Update $\underline{Q}$. We update $\underline{Q}$ by minimizing L'_ρ with other variables fixed as follows

$$\begin{aligned}
& \min_{\underline{Q}^\top * \underline{Q} = \underline{I}_r} L_\rho(\underline{L}^{t+1}, \underline{S}^{t+1}, \underline{Q}, \underline{X}^t, \underline{R}^t, \underline{Y}_1^t, \underline{Y}_2^t) \\
&= \min_{\underline{Q}^\top * \underline{Q} = \underline{I}_r} \left\langle \underline{Y}_1^t, \underline{Q} * \underline{X}^t - \underline{L}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{Q} * \underline{X}^t - \underline{L}^{t+1}\|_F^2. \\
&= \min_{\underline{Q}^\top * \underline{Q} = \underline{I}_r} \frac{\rho}{2} \|\underline{Q} * \underline{X}^t - (\underline{L}^{t+1} - \rho^{-1} \underline{Y}_1^t)\|_F^2 \\
&= \mathfrak{P} \left((\underline{L}^{t+1} - \rho^{-1} \underline{Y}_1^t) * (\underline{X}^t)^\top \right),
\end{aligned} \tag{45}$$

where operator $\mathfrak{P}(\cdot)$ is defined in Lemma 5 as follows.

Lemma 5 ([51]). *Given any tensors $\underline{A} \in \mathbb{R}^{r \times n_2 \times n_3}$, $\underline{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, suppose tensor $\underline{B} * \underline{A}^\top$ has t -SVD $\underline{B} * \underline{A}^\top = \underline{U} * \underline{\Lambda} * \underline{V}^\top$, where $\underline{U} \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\underline{V} \in \mathbb{R}^{r \times r \times n_3}$. Then, the problem*

$$\min_{\underline{Q}^\top * \underline{Q} = \underline{I}_r} \|\underline{P} * \underline{A} - \underline{B}\|_F^2 \tag{46}$$

has a closed-form solution as

$$\underline{Q} = \mathfrak{P}(\underline{B} * \underline{A}^\top) := \underline{U} * \underline{V}^\top. \tag{47}$$

- Update $(\underline{X}, \underline{R})$: We update $(\underline{X}, \underline{S})$ by minimizing L'_ρ with other variables fixed as follows

$$\begin{aligned}
& \min_{(\underline{X}, \underline{R})} L_\rho(\underline{L}^{t+1}, \underline{S}^{t+1}, \underline{Q}^{t+1}, \underline{X}, \underline{R}, \underline{Y}_1^t, \underline{Y}_2^t) \\
&= \min_{(\underline{X}, \underline{R})} \gamma \|\underline{X}\|_{\text{TNN}} + \gamma \lambda \|\underline{R}\|_1 + \left\langle \underline{Y}_1^t, \underline{Q}^{t+1} * \underline{X} - \underline{L}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{Q}^{t+1} * \underline{X} - \underline{L}^{t+1}\|_F^2 \\
&\quad + \left\langle \underline{Y}_2^t, \underline{R} - \underline{S}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{R} - \underline{S}^{t+1}\|_F^2.
\end{aligned} \tag{48}$$

Please note that Problem (48) can further be solved separately as follows

$$\begin{aligned}
\underline{K}^{t+1} &= \underset{\underline{X}}{\operatorname{argmin}} \gamma \|\underline{X}\|_{\text{TNN}} + \left\langle \underline{Y}_1^t, \underline{Q}^{t+1} * \underline{X} - \underline{L}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{Q}^\top * \underline{X} - \underline{L}^{t+1}\|_F^2 \\
&= \underset{\underline{X}}{\operatorname{argmin}} \gamma \|\underline{X}\|_{\text{TNN}} + \frac{\rho}{2} \|\underline{Q}^{t+1} * \underline{X} - (\underline{L}^{t+1} - \rho^{-1} \underline{Y}_1^t)\|_F^2 \\
&\stackrel{(i)}{=} \underset{\underline{X}}{\operatorname{argmin}} \gamma \|\underline{X}\|_{\text{TNN}} + \frac{\rho}{2} \|\underline{X} - (\underline{Q}^{t+1})^\top * (\underline{L}^{t+1} - \rho^{-1} \underline{Y}_1^t)\|_F^2 \\
&= \mathfrak{S}_{\gamma \rho^{-1}}^{\|\cdot\|_{\text{TNN}}} \left((\underline{Q}^{t+1})^\top * (\underline{L}^{t+1} - \rho^{-1} \underline{Y}_1^t) \right)
\end{aligned} \tag{49}$$

and

$$\begin{aligned}
\underline{R}^{t+1} &= \underset{\underline{R}}{\operatorname{argmin}} \gamma \lambda \|\underline{R}\|_1 + \left\langle \underline{Y}_2^t, \underline{R} - \underline{S}^{t+1} \right\rangle + \frac{\rho}{2} \|\underline{R} - \underline{S}^{t+1}\|_F^2 \\
&= \mathfrak{S}_{\gamma \lambda \rho^{-1}}^{\|\cdot\|_1} \left(\underline{K}^{t+1} - \rho^{-1} \underline{Y}_2^t \right).
\end{aligned} \tag{50}$$

The equality (i) in Equation (49) holds because according to $\underline{Q}^\top * \underline{Q} = \underline{I}$, we have

$$\begin{aligned}
\min_{\underline{X}} \|\underline{Q} * \underline{X} - \underline{Y}\|_F^2 &= \min_{\underline{X}} \frac{1}{n_3} \|\overline{\underline{Q}} \cdot \overline{\underline{X}} - \overline{\underline{Y}}\|_F^2 \\
&= \min_{\underline{X}} \frac{1}{n_3} \|\overline{\underline{Y}}\|_F^2 - \frac{2}{n_3} \langle \overline{\underline{Q}} \cdot \overline{\underline{X}}, \overline{\underline{Y}} \rangle + \frac{1}{n_3} \|\overline{\underline{Q}} \cdot \overline{\underline{X}}\|_F^2 \\
&= \min_{\underline{X}} \frac{1}{n_3} \|\overline{\underline{Y}}\|_F^2 - \frac{2}{n_3} \langle \overline{\underline{X}}, \overline{\underline{Q}}^H \overline{\underline{Y}} \rangle + \frac{1}{n_3} \|\overline{\underline{X}}\|_F^2 \\
&= \min_{\underline{X}} \frac{1}{n_3} \|\overline{\underline{X}} - \overline{\underline{Q}}^H \overline{\underline{Y}}\|_F^2 \\
&= \min_{\underline{X}} \frac{\rho}{2} \|\underline{X} - \underline{Q}^\top * \underline{Y}\|_F^2.
\end{aligned} \tag{51}$$

- Update $(\underline{Y}_1, \underline{Y}_2)$. The Lagrangian multipliers are updated by gradient ascent as follows

$$\begin{aligned}
\underline{Y}_1^{t+1} &= \underline{Y}_1^t + \rho(\underline{Q}^{t+1} * \underline{X}^{t+1} - \underline{L}^{t+1}), \\
\underline{Y}_2^{t+1} &= \underline{Y}_2^t + \rho(\underline{R}^{t+1} - \underline{S}^{t+1}).
\end{aligned} \tag{52}$$

Algorithm 2 Solving Problem (40) using ADMM.

Input: The observed tensor $\underline{M}$, an upper estimation r of $r_{\text{tubal}}(\underline{L}_0)$, the parameters $\gamma, \lambda, \rho, \delta$.

1: Initialize $t = 0, \underline{L}^0 = \underline{S}^0 = \underline{R}^0 = \underline{Y}_1^0 = \underline{Y}_2^0 = \underline{0} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \underline{Q}^0 = \underline{0} \in \mathbb{R}^{n_1 \times r \times n_3}, \underline{X}^0 = \underline{0} \in \mathbb{R}^{r \times n_2 \times n_3}$.

2: **for** $t = 0, \dots, T_{\max}$ **do**

3: Update $(\underline{L}^{t+1}, \underline{S}^{t+1})$ by Equation (42);

4: Update $\underline{Q}^{t+1}$ by Equation (45);

5: Update $(\underline{X}^{t+1}, \underline{R}^{t+1})$ by Equations (49)–(50);

6: Update $(\underline{Y}_1^{t+1}, \underline{Y}_2^{t+1})$ by Equation (52);

7: Check the convergence criteria:

(i) convergence of variables: $\|\underline{A}^{t+1} - \underline{A}^t\|_\infty \leq \delta, \forall \underline{A} \in \{\underline{L}, \underline{S}, \underline{R}, \underline{Q}, \underline{X}\}$

(ii) convergence of constraints: $\max\{\|\underline{Q}^{t+1} * \underline{X}^{t+1} - \underline{L}^t\|_\infty, \|\underline{R}^{t+1} - \underline{S}^{t+1}\|_\infty\} \leq \delta$.

8: **end for**

Output: $(\hat{\underline{L}}, \hat{\underline{S}}) = (\underline{L}^{t+1}, \underline{S}^{t+1})$.

The algorithmic steps are summarized in Algorithm 2. The complexity analysis is given as follows.

In each iteration of Algorithm 2, the update of $\underline{L}$ requires FFT/IFFT, and n_3 multiplications of n_1 -by- r and r -by- n_2 matrices, which costs $O((n_1 n_2 + r n_1 + r n_2) n_3 \log n_3 + r n_1 n_2 n_3)$; updating $\underline{S}$ costs $O(n_1 n_2 n_3)$; updating of $\underline{Q}$ involves FFT/IFFT and n_3 SVDs of n_1 -by- r matrices, which costs $O(r n_1 n_3 \log n_3 + r^2 n_1 n_3)$; updating $\underline{X}$ involves FFT/IFFT and n_3 SVDs of r -by- n_2 , which costs $O(r n_2 n_3 \log n_3 + r^2 n_2 n_3)$. Then, the per-iteration computational complexity of Algorithm 2 is dominated by

$$O\left(\max\left\{n_1 n_2 n_3 \log n_3, r^2(n_1 + n_2) n_3\right\}\right).$$

Since the low-tubal-rank assumption $r \ll \min\{n_1, n_2\}$ is adopted in this paper, the per-iteration of Algorithm 2 is much lower than Algorithm 1.

5. Experiments

5.1. Synthetic Data

We first verify the correctness of Theorem 1. Specifically, we check whether the following two statements indicated in Theorem 1 hold in experiments on synthetic data sets:

- (I). (Exact recovery in the noiseless setting.) Our analysis guarantees that the underlying low-rank tensor $\underline{L}_0$ and sparse tensor $\underline{S}_0$ can be exactly recovered in the noiseless setting. This statement will be checked in Section 5.1.1.
- (II). (Linear scaling of errors with the noise level.) In Theorem 1, the estimation errors on $\underline{L}_0$ and $\underline{S}_0$ scales linearly with the noise level δ . This statement will be checked in Section 5.1.2.

Signal Generation. With a given tubal rank r_0 , we first generate the underlying tensor $\underline{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ by $\underline{L}_0 = \underline{A} * \underline{B} / n_3$, where tensors $\underline{A} \in \mathbb{R}^{n_1 \times r_0 \times n_3}$ and $\underline{B} \in \mathbb{R}^{r_0 \times n_2 \times n_3}$ are generated with *i.i.d.* standard Gaussian elements. Then, the sparse corruption tensor $\underline{S}_0$ is generated by choosing its support uniformly at random. The non-zero elements of $\underline{S}_0$ will be *i.i.d.* sampled from a certain distribution that will be specified afterwards. Furthermore, the noise tensor $\underline{E}_0$ is generated with entries sampled *i.i.d.* from $\mathcal{N}(0, \sigma^2)$ with $\sigma = c \|\underline{L}_0\|_F / \sqrt{n_1 n_2 n_3}$, where we set constant c is to control the signal noise ratio. Finally, the observed tensor $\underline{M}$ is formed by $\underline{M} = \underline{L}_0 + \underline{S}_0 + \underline{E}_0$.

5.1.1. Exact Recovery in the Noiseless Setting

We first check *Statement (I)*, i.e., exact recovery in the noiseless setting. Specifically, we will show that Algorithm 1 and Algorithm 2 can exactly recover the underlying tensor $\underline{L}_0$ and the sparse corruption $\underline{S}_0$. We first test the recovery performance of different tensor sizes by setting $n = n_1 = n_2 \in \{100, 160, 200\}$ and $n_3 = 20$, with $(r_{\text{tubal}}(\underline{L}_0), \|\underline{S}_0\|_0) = (0.05n, 0.05n^2n_3)$. The non-zero elements of tensor $\underline{S}_0$ is sampled from *i.i.d.* symmetric Bernoulli distribution, i.e., the possibility of being 1 or -1 are 1/2. The results are shown in Table 1. It can be seen that both Algorithm 1 and Algorithm 2 can obtain relative standard error (RSE) smaller than $1e-5$ by which we can say that $\underline{L}_0$ and $\underline{S}_0$ are exact recovered. We can also see that Algorithm 2 runs much faster than Algorithm 1.

Table 1. Performance of Algorithm 1 and Algorithm 2 in both accuracy and speed for different tensor sizes when the gross corruption. Outliers from symmetric Bernoulli, observation tensor $\underline{M} \in \mathbb{R}^{n \times n \times n_3}$, $n_3 = 30$, $r_{\text{tubal}}(\underline{L}_0) = 0.05n$, $\|\underline{S}_0\|_1 = 0.05n^2n_3$, noise level $c = 0$, $r = \max\{\lfloor 2r_{\text{tubal}}(\underline{L}_0) \rfloor, 15\}$.

n	$r_{\text{tubal}}(\underline{L}_0)$	$\ \underline{S}_0\ _0$	Method	$r_{\text{tubal}}(\hat{\underline{L}})$	$\frac{\ \hat{\underline{L}} - \underline{L}_0\ _F}{\ \underline{L}_0\ _F}$	$\frac{\ \hat{\underline{S}} - \underline{S}_0\ _F}{\ \underline{S}_0\ _F}$	time/s
100	5	1×10^4	Algorithm 1	5	5.13×10^{-6}	5.27×10^{-6}	3.63
			Algorithm 2	5	4.92×10^{-6}	5.12×10^{-6}	1.76
160	8	2.56×10^4	Algorithm 1	8	3.86×10^{-6}	3.52×10^{-6}	9.52
			Algorithm 2	8	4.48×10^{-6}	4.08×10^{-6}	4.42
200	10	4×10^4	Algorithm 1	10	3.46×10^{-6}	3.59×10^{-6}	14.16
			Algorithm 2	10	4.12×10^{-6}	4.63×10^{-6}	7.44

We then test whether the recovery performance can be affected by the distribution of the corruptions. This is done by choosing the non-zeros elements of $\underline{S}_0$ from *i.i.d.* standard Gaussian distribution. The experimental results are reported in Table 2. We can find that both Algorithm 1 and Algorithm 2 can exactly recover the true $\underline{L}_0$ and $\underline{S}_0$ and Algorithm 2 runs much faster than Algorithm 1.

We also conduct STPCP by Algorithm 1 and Algorithm 2 with missing entries. After generating $\underline{L}_0$, $\underline{S}_0$ and $\underline{E}_0$, we get the observation by Model (27). We choose the support of $\underline{B}$ uniformly at random with possibility 0.8 and then set elements in the chosen support to be 1. Thus, %20 of the entries are missing. The corrupted observation $\underline{M}$ is then formed by $\underline{M} = \underline{B} \odot (\underline{L}_0 + \underline{S}_0 + \underline{E}_0)$. We show the recover results in Table 3. We can see that the underlying low-rank tensor $\underline{L}_0$ can be exactly recovered and the observed part of the corruption tensor $\underline{B} \odot \underline{S}_0$ can also be exactly recovered (Please note that it is impossible to recover the unobserved entries of a sparse tensor $\underline{S}_0$ [52]).

Table 2. Performance of Algorithm 1 and Algorithm 2 in both accuracy and speed for different tensor sizes when the gross corruption. Outliers from standard Gaussian distribution, observation tensor $\underline{\mathbf{M}} \in \mathbb{R}^{n \times n \times n_3}$, $n_3 = 30$, $r_{\text{tubal}}(\underline{\mathbf{L}}_0) = 0.05n$, $\|\underline{\mathbf{S}}_0\|_1 = 0.05n^2n_3$, noise level $c = 0$, $r = \max\{\lfloor 2r_{\text{tubal}}(\underline{\mathbf{L}}_0) \rfloor, 15\}$.

n	$r_{\text{tubal}}(\underline{\mathbf{L}}_0)$	$\ \underline{\mathbf{S}}_0\ _0$	Method	$r_{\text{tubal}}(\hat{\underline{\mathbf{L}}})$	$\frac{\ \hat{\underline{\mathbf{L}}} - \underline{\mathbf{L}}_0\ _F}{\ \underline{\mathbf{L}}_0\ _F}$	$\frac{\ \hat{\underline{\mathbf{S}}} - \underline{\mathbf{S}}_0\ _F}{\ \underline{\mathbf{S}}_0\ _F}$	time/s
100	5	1×10^4	Algorithm 1	5	2.7×10^{-6}	2.6×10^{-6}	4.43
			Algorithm 2	5	2.9×10^{-6}	3.2×10^{-6}	1.82
160	8	2.56×10^4	Algorithm 1	8	4.76×10^{-6}	4.08×10^{-6}	10.45
			Algorithm 2	8	4.24×10^{-6}	4.05×10^{-6}	5.15
200	10	4×10^4	Algorithm 1	10	3.78×10^{-6}	3.64×10^{-6}	18.97
			Algorithm 2	10	3.78×10^{-6}	3.63×10^{-6}	8.04

Table 3. Performance of Algorithm 1 and Algorithm 2 in both accuracy and speed for different tensor sizes when the gross corruption. Outliers from symmetric Bernoulli, observation tensor $\underline{\mathbf{M}} \in \mathbb{R}^{n \times n \times n_3}$, $n_3 = 30$, $r_{\text{tubal}}(\underline{\mathbf{L}}_0) = 0.05n$, $\|\underline{\mathbf{S}}_0\|_1 = 0.05n^2n_3$, noise level $c = 0$, $r = \max\{\lfloor 2r_{\text{tubal}}(\underline{\mathbf{L}}_0) \rfloor, 15\}$, with %20 random missing entries.

n	$r_{\text{tubal}}(\underline{\mathbf{L}}_0)$	$\ \underline{\mathbf{B}} \odot \underline{\mathbf{S}}_0\ _0$	Method	$r_{\text{tubal}}(\hat{\underline{\mathbf{L}}})$	$\frac{\ \hat{\underline{\mathbf{L}}} - \underline{\mathbf{L}}_0\ _F}{\ \underline{\mathbf{L}}_0\ _F}$	$\frac{\ \hat{\underline{\mathbf{S}}} - \underline{\mathbf{B}} \odot \underline{\mathbf{S}}_0\ _F}{\ \underline{\mathbf{B}} \odot \underline{\mathbf{S}}_0\ _F}$	time/s
100	5	8×10^3	Algorithm 1	5	7.52×10^{-6}	5.97×10^{-6}	3.87
			Algorithm 2	5	7.50×10^{-6}	5.96×10^{-6}	1.69
160	8	2.048×10^4	Algorithm 1	8	4.46×10^{-6}	5.17×10^{-6}	9.64
			Algorithm 2	8	5.60×10^{-6}	4.71×10^{-6}	4.46
200	10	3.2×10^4	Algorithm 1	10	4.78×10^{-6}	4.04×10^{-6}	14.78
			Algorithm 2	10	5.13×10^{-6}	4.20×10^{-6}	7.77

5.1.2. Linear Scaling of Errors with the Noise Level

We then verify *Statement (II)* that the estimation errors have linear scale behavior with respect to the noise level. The estimation errors are measured using the mean-squared-error (MSE)

$$\text{MSE}(\hat{\underline{\mathbf{L}}}) = \frac{\|\hat{\underline{\mathbf{L}}} - \underline{\mathbf{L}}_0\|_F^2}{n_1 n_2 n_3}, \quad \text{MSE}(\hat{\underline{\mathbf{S}}}) = \frac{\|\hat{\underline{\mathbf{S}}} - \underline{\mathbf{S}}_0\|_F^2}{n_1 n_2 n_3},$$

for the low rank component $\underline{\mathbf{L}}_0$ and the sparse component $\underline{\mathbf{S}}_0$, respectively. We test tensors of 3 different size by choosing $n \in \{60, 80, 100\}$ and $n_3 = 20$. The tubal rank $r_{\text{tubal}}(\underline{\mathbf{L}}_0)$ of $\underline{\mathbf{L}}_0$ and sparsity s of $\underline{\mathbf{S}}_0$ are set as $(r_{\text{tubal}}(\underline{\mathbf{L}}_0), s) = (5, 0.1n^2n_3)$. We vary the signal noise ratio $c = 0.03 : 0.03 : 0.6$ which is in proportional of the noise level δ . We run the proposed Algorithm 1, test 50 trials, and report the averaged MSEs. The MSEs of $\hat{\underline{\mathbf{L}}}$ and $\hat{\underline{\mathbf{S}}}$ versus c^2 are shown in sub-figures (a) and (b) in Figure 4. We can see that the estimation error has linear scaling behavior along with the noise level as Theorem 1 indicates. Since the results for $n = 80$ and $n = 100$ are quite similar to the case of $n = 60$, they are simply omitted.

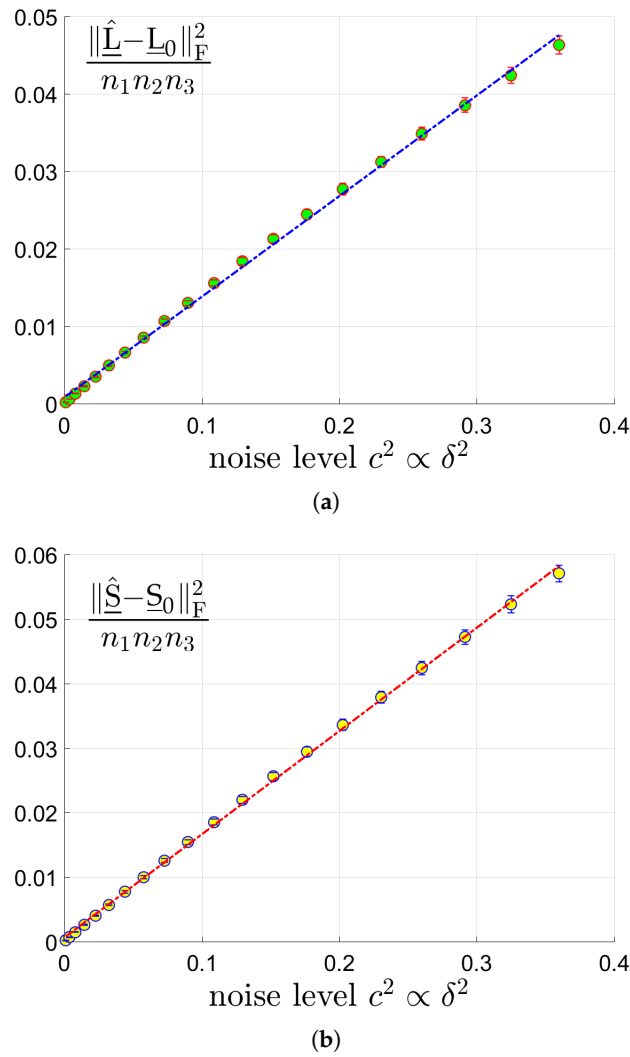


Figure 4. The MSEs of $\hat{\underline{\mathbf{L}}}$ and $\hat{\underline{\mathbf{S}}}$ versus c^2 for tensors of size $60 \times 60 \times 20$ where the tubal rank $r_{\text{tubal}}(\underline{\mathbf{L}}_0)$ of $\underline{\mathbf{L}}_0$ and sparsity s of $\underline{\mathbf{S}}_0$ are set as $(r_{\text{tubal}}(\underline{\mathbf{L}}_0), s) = (5, 0.1n^2n_3)$. (a): MSE of $\hat{\underline{\mathbf{L}}}$ vs c^2 . (b): MSE of $\hat{\underline{\mathbf{S}}}$ vs c^2 .

5.2. Real Data Sets

In this section, experiments on real data sets (color images and videos) are carried out to evaluate the effectiveness and efficiency of the proposed Algorithms 1 and 2. Besides noises and sparse corruptions, we also consider missing values which is more challenging. The proposed algorithms are compared with the following typical models:

- (I). NN-I: tensor recovery based on matrix nuclear norms of frontal slices formulated as follows

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{M}} - \underline{\mathbf{L}} - \underline{\mathbf{E}})\|_F + \gamma \sum_{k=1}^{n_3} (\|\mathbf{L}_{(k)}\|_* + \lambda \|\mathbf{S}^{(k)}\|_1). \quad (53)$$

This model will be used for image restoration in Section 5.2.1. Please note that Model (53) is equivalent to parallel matrix recovery on each frontal slice.

- (II). NN-II: tensor recovery based on matrix nuclear norm formulated as follows

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{S}}} \frac{1}{2} \|\underline{\mathbf{B}} \odot (\underline{\mathbf{M}} - \underline{\mathbf{L}} - \underline{\mathbf{E}})\|_F + \gamma \|\underline{\mathbf{L}}\|_* + \gamma \lambda \|\underline{\mathbf{S}}\|_1, \quad (54)$$

where $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_{n_3}] \in \mathbb{R}^{n_1 n_2 \times n_3}$ with $\mathbf{l}_k := \text{vec}(\mathbf{L}^{(k)}) \in \mathbb{R}^{n_1 n_2}$ defined as the vectorization [40] of frontal slices $\mathbf{L}^{(k)}$, for all $k = 1, 2, \dots, n_3$. This model will be used for video restoration in Section 5.2.2.

(III). SNN: tensor recovery based on SNN formulated as follows

$$\min_{\mathbf{L}, \mathbf{S}} \frac{1}{2} \|\mathbf{B} \odot (\mathbf{M} - \mathbf{L} - \mathbf{E})\|_F + \gamma \sum_{i=1}^3 \alpha_m \|\mathbf{L}_{(i)}\|_* + \gamma \|\mathbf{S}\|_1, \quad (55)$$

where $\mathbf{L}_{(i)} \in \mathbb{R}^{n_i \times \prod_{j \neq i} n_j}$ is the mode- i matriculation of tensor $\mathbf{L} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, for all $i = 1, 2, 3$.

We solve the above Model (53)–(55) using ADMM implemented by ourselves in Matlab. The effectiveness of the algorithms is measured by Peaks Signal Noise Ratio (PSNR)

$$\text{PSNR} := 10 \log_{10} \left(\frac{n_1 n_2 n_3 \|\mathbf{L}_0\|_\infty^2}{\|\hat{\mathbf{L}} - \mathbf{L}_0\|_F^2} \right),$$

Please note that a larger PSNR value indicates higher quality of $\hat{\mathbf{L}}$.

5.2.1. Color Images

Color images are the most commonly used 3-way tensors. We test the twenty 256-by-256-by-3 color images which have been used in [37], and carry out robust tensor recovery with missing entries (see Figure 5). Following [37], for a color image $\mathbf{L}_0 \in \mathbb{R}^{n \times n \times 3}$, we choose its support uniformly at random with ratio ρ_s and fill in the values with *i.i.d.* symmetric Bernoulli variables to generate $\mathbf{S}_0$. The noise tensor $\mathbf{E}_0$ is generated with *i.i.d.* zero-mean Gaussian entries whose standard deviation is given by $\sigma = 0.05 \|\mathbf{L}_0\|_F / \sqrt{3n^2}$. Then, we form the binary observation mask $\mathbf{B}$ by choosing its support uniformly at random with ratio ρ_{obs} . Finally, the partially observed corruption $\mathbf{M} = \mathbf{B} \odot (\mathbf{L}_0 + \mathbf{S}_0 + \mathbf{E}_0)$ are formed.

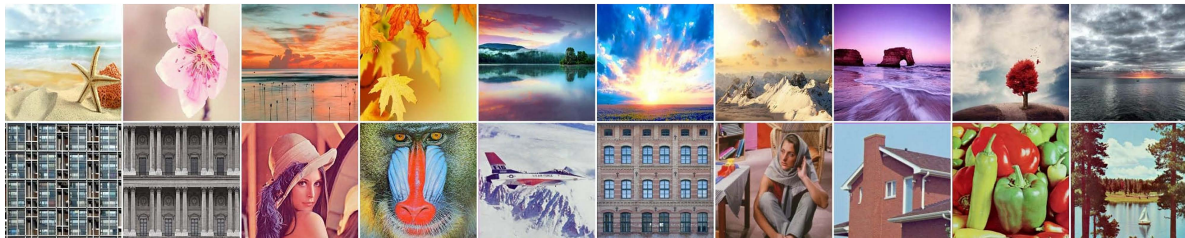


Figure 5. The 20 color images used.

We consider two scenarios by setting $(\rho_{\text{obs}}, \rho_s) \in \{(0.9, 0.1), (0.8, 0.2)\}$. For NN (Model (53)), we set the regularization parameters $\lambda = 1 / \sqrt{n \rho_{\text{obs}}}$ (suggested by [46]), and set parameter $\gamma = \|\mathbf{E}_0\|_{\text{sp}}$ where $\|\mathbf{E}_0\|_{\text{sp}}$ is estimated as $6.5\sigma \sqrt{3\rho_{\text{obs}} n \log(6n)}$ (suggested by [5]). For SNN, the parameters are chosen to satisfy $\gamma = 0.05$, $\alpha_1 = \alpha_2 = \sqrt{3n\rho_{\text{obs}}}$, $\alpha_3 = 0.01 \sqrt{3n\rho_{\text{obs}}}$. For Algorithm 1 and Algorithm 2, we set $\gamma = 0.3 \|\mathbf{E}_0\|_{\text{sp}}$, and $\lambda = 1 / \sqrt{3n\rho_{\text{obs}}}$. The initialized rank r in Algorithm 2 is set as 60. In each setting, we test each color image for 10 times and report the averaged PSNR and time. For quantitative comparison, we show the PSNR values and running times in Figures 6 and 7 for settings of $(\rho_{\text{obs}}, \rho_s) = (0.9, 0.1)$ and $(\rho_{\text{obs}}, \rho_s) = (0.8, 0.2)$, respectively. Several visual examples are shown in Figure 8 for qualitative comparison for the setting of $(\rho_{\text{obs}}, \rho_s) = (0.8, 0.2)$. We can see from Figures 6–8 that the proposed Algorithm 1 has the highest recovery quality and the proposed Algorithm 2 has the second highest quality but the fastest running time.

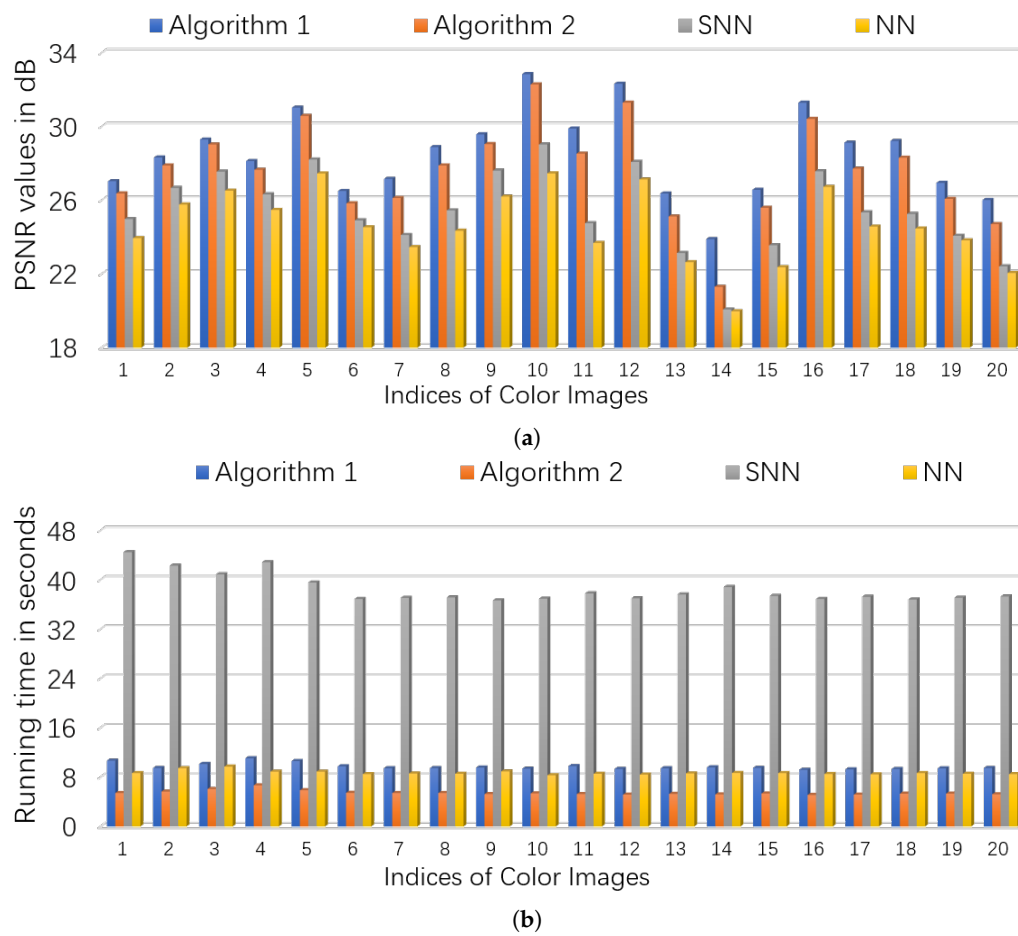


Figure 6. The quantitative comparison in PSNR and time on color images. First, 10% entries of each image is corrupted by *i.i.d.* symmetric Bernoulli variable, then polluted by Gaussian noise of noise level $c = 0.05$, and finally 10% of the corrupted entries are missing uniformly at random. (a): the PSNR values of each algorithm; (b): the running time of each algorithm.

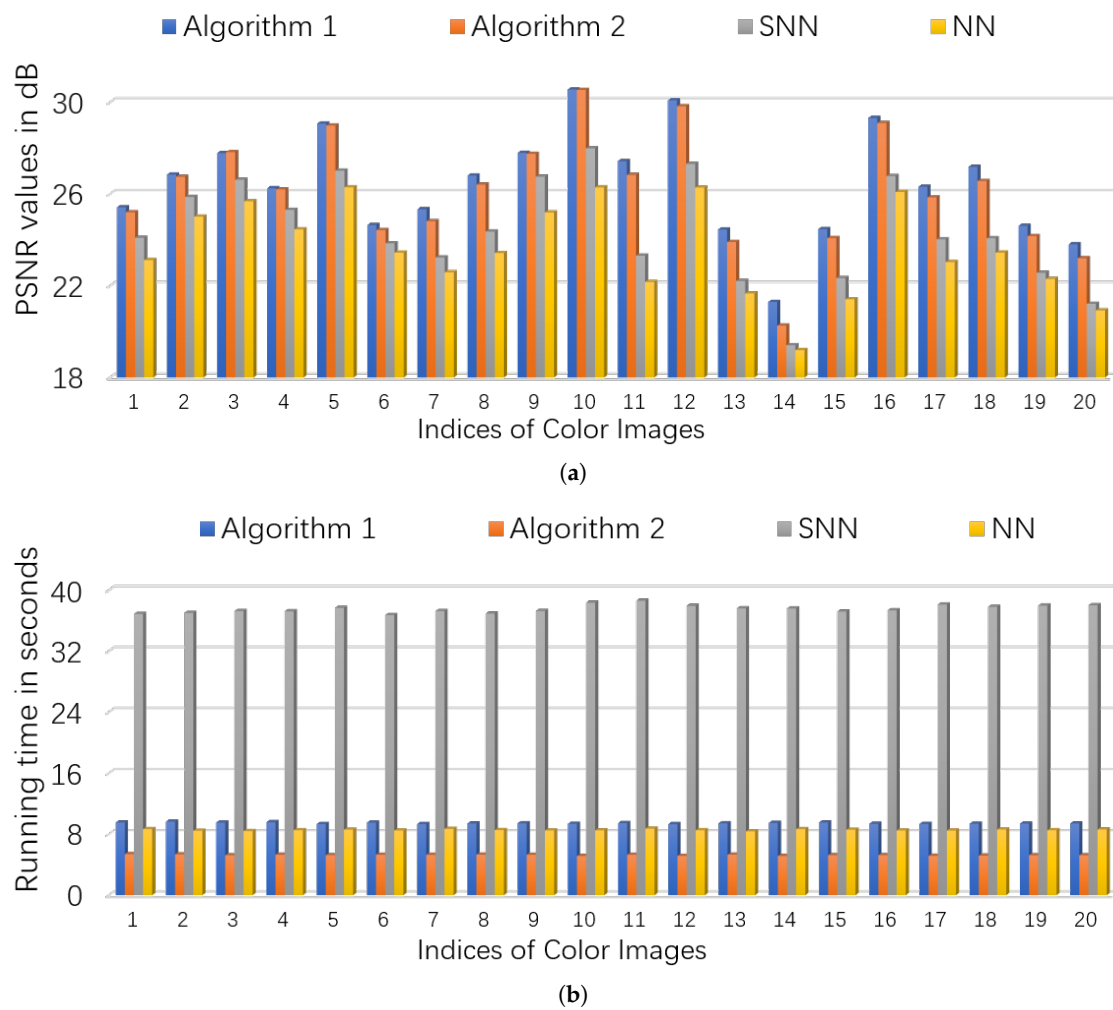


Figure 7. The quantitative comparison in PSNR and time on color images. First, 20% entries of each image is corrupted by *i.i.d.* symmetric Bernoulli variable, then polluted by Gaussian noise of noise level $c = 0.05$, and finally 20% of the corrupted entries are missing uniformly at random. (a): the PSNR values of each algorithm; (b): the running time of each algorithm.

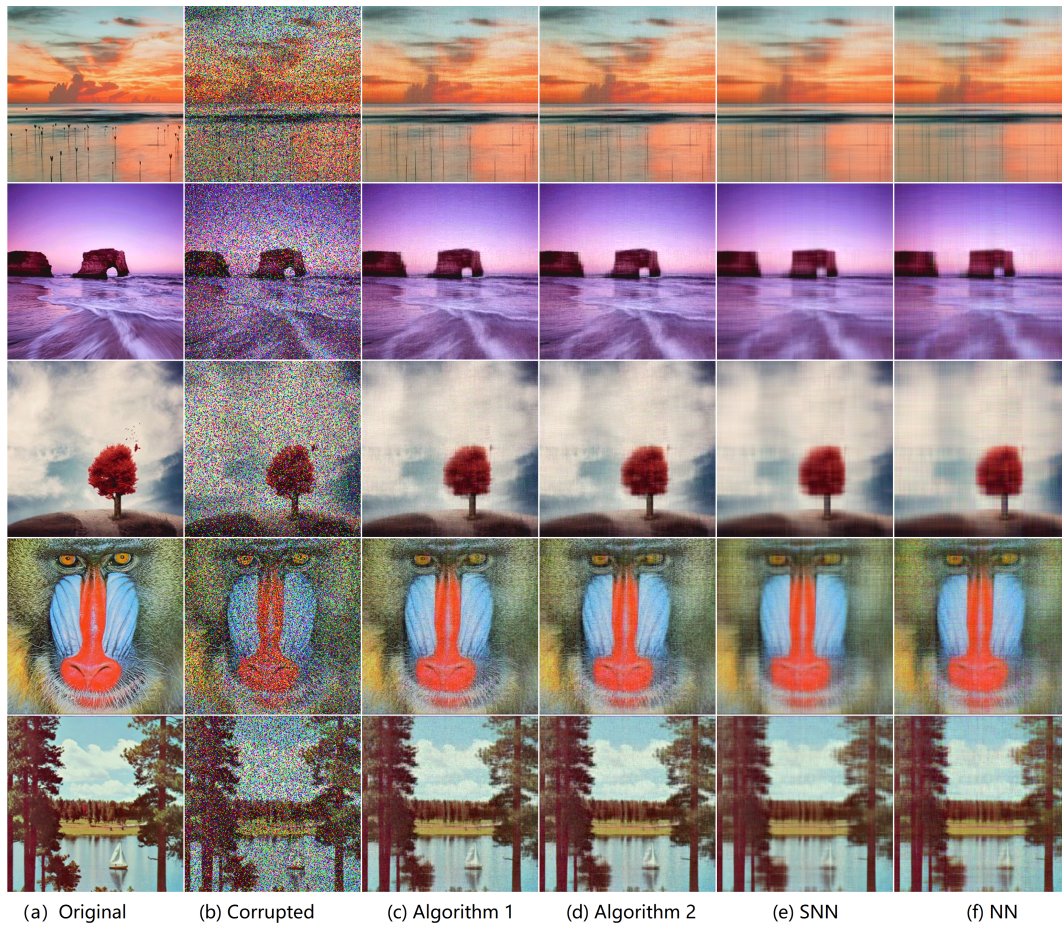


Figure 8. The visual results for image recovery of different algorithms. First, 20% entries of each image is corrupted by *i.i.d.* symmetric Bernoulli variable, then polluted by Gaussian noise of noise level $c = 0.05$, and finally 20% of the corrupted entries are missing uniformly at random. (a): the original image; (b): the corrupted image; (c) image recovered by Algorithm 1; (d) image recovered by Algorithm 2; (e) image recovered by the matrix nuclear norm (NN)-based Model (53); (f) image recovered by the SNN-based Model (55).

5.2.2. Videos

In this subsection, video restoration experiments are conducted on four broadly used YUV videos (They can be downloaded from <https://sites.google.com/site/subudhibadri/fewhelpfuldownloads>: Akiyo_qcif, Scilnet_qcif, Carphone_qcif, and Claire_qcif.) Owing to computational limitation, we simply use the first 32 frames of the Y components of all the videos which results in four 144-by-176-by-30 tensors. For a 3-way data tensor $\underline{\mathbf{L}}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, To generate corruption $\underline{\mathbf{S}}_0$, the support is chosen uniformly at random with ratio ρ_s and then elements in the support are filled in with *i.i.d.* symmetric Bernoulli variables. The noise tensor $\underline{\mathbf{E}}_0$ is also generated with *i.i.d.* zero-mean Gaussian entries whose standard deviation is given by $\sigma = 0.05 \|\underline{\mathbf{L}}_0\|_F / \sqrt{n_1 n_2 n_3}$. Then, the binary observation mask $\underline{\mathbf{B}}$ is formed thorough choosing its support uniformly at random with ratio ρ_{obs} . Finally, the partially observed corruption $\underline{\mathbf{M}} = \underline{\mathbf{B}} \odot (\underline{\mathbf{L}}_0 + \underline{\mathbf{S}}_0 + \underline{\mathbf{E}}_0)$ are formed.

We also consider two scenarios by setting $(\rho_{obs}, \rho_s) \in \{(0.9, 0.1), (0.8, 0.2)\}$. NN-II Model (54) is used in video restoration. For NN-II, we set the regularization parameters $\lambda = 1 / \sqrt{n_1 n_2 \rho_{obs}}$ (suggested by [46]), and set parameter $\gamma = \|\underline{\mathbf{E}}_0\|_{sp}$ where $\|\underline{\mathbf{E}}_0\|_{sp}$ is estimated as $6.5\sigma \sqrt{\rho_{obs} n_1 n_3 \log((n_1 + n_2) n_3)}$ (suggested by [5]). For SNN, the parameters are chosen to satisfy $\gamma = 0.05, \alpha_1 = \alpha_2 = \sqrt{n_1 n_3 \rho_{obs}}, \alpha_3 = 5\sqrt{n_1 n_3 \rho_{obs}}$. For Algorithm 1 and Algorithm 2, we set $\gamma = 0.3 \|\underline{\mathbf{E}}_0\|_{sp}$, and $\lambda = 1 / \sqrt{\max\{n_1, n_2\} n_3 \rho_{obs}}$ after careful parameter tuning. The initialized rank r in Algorithm 2 is set as 60. In each setting, we test

each video for 10 times and report the averaged PSNR and time. For quantitative comparison, we show the PSNR values and running times in Table 4. It can be seen that Algorithm 1 has the highest recovery quality and the proposed Algorithm 2 has the second highest quality but the fastest running time.

Table 4. PSNR values and running time (in seconds) of different algorithms on video data. First, $\rho_s n_1 n_2 n_3$ entries of each image is corrupted by *i.i.d.* symmetric Bernoulli variable, then polluted by Gaussian noise of noise level $c = 0.05$, and finally $(1 - \rho_{\text{obs}}) n_1 n_2 n_3$ of the corrupted entries are missing uniformly at random. The items with **highest PSNR values** are highlighted with bold face, and the items with shortest running time are highlighted with underline.

Data Set	$(\rho_{\text{obs}}, \rho_s)$	Index	NN, Model (54)	SNN, Model (55)	Algorithm 1	Algorithm 2
Akiyo	(0.9,0.1)	PSNR	31.74	32.09	33.94	33.36
		time/s	29.48	51.13	20.10	<u>12.39</u>
	(0.8,0.2)	PSNR	30.59	30.70	32.44	32.07
		time/s	30.65	51.17	19.53	<u>14.92</u>
Silent	(0.9,0.1)	PSNR	28.26	30.39	31.74	31.23
		time/s	28.91	49.79	21.21	<u>14.76</u>
	(0.8,0.2)	PSNR	26.95	27.60	30.42	30.07
		time/s	36.51	60.81	22.43	<u>15.62</u>
Carphone	(0.9,0.1)	PSNR	26.87	28.79	29.15	28.94
		time/s	28.55	47.17	22.12	<u>14.41</u>
	(0.8,0.2)	PSNR	26.12	26.43	28.17	27.99
		time/s	26.72	49.21	20.55	<u>14.74</u>
Claire	(0.9,0.1)	PSNR	30.56	32.20	34.27	34.02
		time/s	29.75	47.32	21.43	<u>13.52</u>
	(0.8,0.2)	PSNR	29.94	30.43	32.96	32.78
		time/s	29.43	50.46	19.47	<u>13.04</u>

6. Conclusions

This paper studied the problem of stable tensor principal component pursuit which aims to recover a tensor from noises and sparse corruptions. We proposed a constrained tubal nuclear norm-based model and established upper bounds on the estimation error. In contrast to prior work [37], our theory can guarantee exact recovery in the noiseless setting. We also designed two algorithms, the first ADMM algorithm can be accelerated by the second Algorithm which adopts a factorization strategy. We validated the correctness of our theory by simulations on synthetic data, and evaluated the effectiveness and efficiency of the proposed algorithms via experiments on color images and videos.

For future directions, it is a natural and interesting extension to consider recovery of 4-way tensors [35] with arbitrary linear transformation [53,54]. It is also interesting to use tensor factorization-based methods [55,56] for STPCP. Another challenging future direction is developing tools to verify whether the unknown tensor satisfies the tensor incoherence condition from its incomplete or corrupted observations.

For extensions of the proposed approach to higher-way tensors, we produce the following two ideas:

1. By recursively applying DFT over successive modes higher than 3 and then unfolding the obtained tensor into 3-way [57], the proposed algorithms and theoretical analysis can be extended to higher-way tensors.
2. By using the overlapped orientation invariant tubal nuclear norm [58], we can extend the proposed algorithm to higher-order cases and obtain orientation invariance.

Author Contributions: Conceptualization, W.F. Data curation, D.W. and R.Z. Formal analysis, W.F. Methodology, W.F., D.W. and R.Z. Software, D.W. and R.Z. Writing, original draft, W.F., D.W. and R.Z.

Funding: This research was funded by the Key Projects of Natural Science Research in Universities in Anhui Province under grant number KJ2019A0994.

Acknowledgments: We sincerely thank Andong Wang who shared the codes of [37] and gave us some suggestions of the proof.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs of Lemmas and Theorems

Appendix A.1. The Proof of Theorem 1

Appendix A.1.1. Key Lemmas for the Proof of Theorem 1

Before Proving Theorem 1, we should define some notations and operators first.

Suppose $\underline{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r has the skinny t-SVD $\underline{L}_0 = \underline{U} * \underline{A} * \underline{V}^\top$, where $\underline{U} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\underline{V} \in \mathbb{R}^{r \times n_2 \times n_3}$ are orthogonal tensors, and $\underline{A} \in \mathbb{R}^{r \times r \times n_3}$ is an f -diagonal tensor. Define the following set

$$\mathbf{T} := \left\{ \underline{U} * \underline{A} + \underline{B} * \underline{V}^\top \mid \underline{A} \in \mathbb{R}^{r \times n_2 \times n_3}, \underline{B} \in \mathbb{R}^{n_1 \times r \times n_3} \right\} \subset \mathbb{R}^{n_1 \times n_2 \times n_3}. \quad (\text{A1})$$

Then, define the projector onto $\mathbf{T}$ for any tensor $\underline{I} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as follows

$$\begin{aligned} \mathcal{P}_{\mathbf{T}}(\underline{I}) &:= \underline{U} * \underline{U}^\top * \underline{I} + \underline{I} * \underline{V} * \underline{V}^\top - \underline{U} * \underline{U}^\top * \underline{I} * \underline{V} * \underline{V}^\top, \\ \mathcal{P}_{\mathbf{T}^\perp}(\underline{I}) &:= (\underline{I} - \underline{U} * \underline{U}^\top) * \underline{I} * (\underline{I} - \underline{V} * \underline{V}^\top). \end{aligned} \quad (\text{A2})$$

Let $\Omega^\perp$ be the complement of $\Omega \subset [n_1] \times [n_2] \times [n_3]$ which is the support of $\underline{S}_0$. Then, define two operators $\mathcal{P}_\Omega, \mathcal{P}_{\Omega^\perp}$ as follows

$$\mathcal{P}_\Omega(\underline{I}) := \sum_{(i,j,k) \in \Omega} \langle \underline{I}, \mathbf{e}_i * \mathbf{e}_k * \mathbf{e}_j^\top \rangle, \quad \mathcal{P}_{\Omega^\perp}(\underline{I}) := \sum_{(i,j,k) \in \Omega^\perp} \langle \underline{I}, \mathbf{e}_i * \mathbf{e}_k * \mathbf{e}_j^\top \rangle, \quad (\text{A3})$$

for any $\underline{I} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

Define two sets $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^\perp$ as follows:

$$\mathbf{\Gamma} = \{(\underline{A}, \underline{A}) \mid \underline{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}\}, \quad \mathbf{\Gamma}^\perp = \{(\underline{A}, -\underline{A}) \mid \underline{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}\}. \quad (\text{A4})$$

Then, for any tensors $\underline{X}_l, \underline{X}_s \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the projectors of the tensor $\underline{X} = (\underline{X}_l, \underline{X}_s)$ into the sets $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^\perp$ are given as follows, respectively

$$\mathcal{P}_{\mathbf{\Gamma}}(\underline{X}) = \left(\frac{\underline{X}_l + \underline{X}_s}{2}, \frac{\underline{X}_l + \underline{X}_s}{2} \right), \quad \mathcal{P}_{\mathbf{\Gamma}^\perp}(\underline{X}) = \left(\frac{\underline{X}_l - \underline{X}_s}{2}, \frac{\underline{X}_s - \underline{X}_l}{2} \right). \quad (\text{A5})$$

For any tensors $\underline{X}_l, \underline{X}_s \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, define two operators on $\underline{X} = (\underline{X}_l, \underline{X}_s)$ as follows:

$$(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})(\underline{X}) = (\mathcal{P}_{\mathbf{T}}(\underline{X}_l), \mathcal{P}_{\Omega}(\underline{X}_s)), \quad (\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(\underline{X}) = (\mathcal{P}_{\mathbf{T}^\perp}(\underline{X}_l), \mathcal{P}_{\Omega^\perp}(\underline{X}_s)). \quad (\text{A6})$$

Also define two norms as follows:

$$\|\underline{X}\|_F = \sqrt{\|\underline{X}_l\|_F^2 + \|\underline{X}_s\|_F^2}, \quad \|\underline{X}\|_{F,\mu} = \sqrt{\|\underline{X}_l\|_F^2 + \mu^2 \|\underline{X}_s\|_F^2}. \quad (\text{A7})$$

where μ is a constant that will be determined afterwards.

We first give Lemma A1 which can be seen as a modified version of Lemma C.1 in [2].

Lemma A1. Assume that $\|\mathcal{P}_\Omega \mathcal{P}_\mathbf{T}\| \leq \frac{1}{2}$, and $\lambda \leq \frac{1}{2\sqrt{n_3}}$. Suppose there exists a tensor $\underline{\mathbf{G}}^*$ satisfying the following conditions

$$\begin{cases} \mathcal{P}_\mathbf{T}(\underline{\mathbf{G}}^*) = \underline{\mathbf{U}} * \underline{\mathbf{V}}^\top, \\ \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\mathbf{G}}^*)\|_{\text{sp}} \leq \frac{1}{2}, \\ \|\mathcal{P}_\Omega(\underline{\mathbf{G}}^* - \lambda \text{sign}(\underline{\mathbf{S}}_0))\|_{\text{F}} \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega^\perp}(\underline{\mathbf{G}}^*)\|_\infty \leq \frac{\lambda}{2}. \end{cases} \quad (\text{A8})$$

Then for any perturbation $\underline{\Delta} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, one has

$$\begin{aligned} & \|\underline{\mathbf{L}}_0 + \underline{\Delta}\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}_0 - \underline{\Delta}\|_1 \\ & \geq \|\underline{\mathbf{L}}_0\|_{\text{TNN}} + \lambda \|\underline{\mathbf{S}}_0\|_1 + \left(\frac{3}{4} - \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\mathbf{G}})\|_{\text{sp}}\right) \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{TNN}} + \left(\frac{3}{4}\lambda - \mathcal{P}_{\Omega^\perp}(\|\underline{\mathbf{G}}\|_\infty)\right) \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1. \end{aligned} \quad (\text{A9})$$

Proof. Let $\underline{\mathbf{G}}_t \in \partial\|\underline{\mathbf{L}}_0\|_{\text{TNN}}$, i.e., any sub-gradient of $\|\cdot\|_{\text{TNN}}$ at $\underline{\mathbf{L}}_0$, then it satisfies:

$$\mathcal{P}_\mathbf{T}(\underline{\mathbf{G}}_t) = \underline{\mathbf{U}} * \underline{\mathbf{V}}^\top, \quad \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\mathbf{G}}_t)\|_{\text{sp}} \leq 1. \quad (\text{A10})$$

$\underline{\mathbf{G}}_t \in \partial\|\underline{\mathbf{L}}_0\|_{\text{TNN}}$ and $\underline{\mathbf{G}}_s \in \partial(\lambda\|\underline{\mathbf{S}}_0\|_1)$. According to the convexity of $\|\cdot\|_{\text{TNN}}$ and $\|\cdot\|_1$, we have

$$\|\underline{\mathbf{L}}_0 + \underline{\Delta}\|_{\text{TNN}} \geq \|\underline{\mathbf{L}}_0\|_{\text{TNN}} + \langle \underline{\mathbf{G}}_t, \underline{\Delta} \rangle, \quad \lambda \|\underline{\mathbf{S}}_0 - \underline{\Delta}\|_1 \geq \lambda \|\underline{\mathbf{S}}_0\|_1 - \langle \underline{\mathbf{G}}_s, \underline{\Delta} \rangle. \quad (\text{A11})$$

By choosing $\underline{\mathbf{G}}_t = \underline{\mathbf{U}} * \underline{\mathbf{V}}^\top + \underline{\mathbf{P}} * \underline{\mathbf{Q}}^\top$, where $\underline{\mathbf{P}}$ and $\underline{\mathbf{Q}}$ comes from the skinny t-SVD of $\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}) = \underline{\mathbf{P}} * \underline{\Sigma} * \underline{\mathbf{Q}}^\top$, one has

$$\begin{aligned} \langle \underline{\mathbf{G}}_t, \underline{\Delta} \rangle &= \langle \underline{\mathbf{G}}, \underline{\Delta} \rangle + \langle \underline{\mathbf{G}}_t - \underline{\mathbf{G}}, \underline{\Delta} \rangle = \langle \underline{\mathbf{G}}, \underline{\Delta} \rangle + \langle \mathcal{P}_{\mathbf{T}^\perp}(\underline{\mathbf{G}}_t), \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}) \rangle - \langle \mathcal{P}_{\mathbf{T}^\perp}(\underline{\mathbf{G}}), \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}) \rangle \\ &= \langle \underline{\mathbf{G}}, \underline{\Delta} \rangle - (1 - \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\mathbf{G}})\|_{\text{sp}}) \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{TNN}}. \end{aligned} \quad (\text{A12})$$

Also, by choosing $\underline{\mathbf{G}}_s = \lambda \text{sign}(\underline{\mathbf{S}}_0) - \text{sign}(\mathcal{P}_{\Omega^\perp}(\underline{\Delta}))$, one has

$$\begin{aligned} -\langle \underline{\mathbf{G}}_s, \underline{\Delta} \rangle &= -\langle \underline{\mathbf{G}}, \underline{\Delta} \rangle - \langle \underline{\mathbf{G}}_s - \underline{\mathbf{G}}, \underline{\Delta} \rangle \\ &= -\langle \underline{\mathbf{G}}, \underline{\Delta} \rangle - \langle \mathcal{P}_\Omega(\lambda \text{sign}(\underline{\mathbf{S}}_0) - \underline{\mathbf{G}}), \mathcal{P}_\Omega(\underline{\Delta}) \rangle - \langle \mathcal{P}_{\Omega^\perp}(\underline{\mathbf{G}}_s), \mathcal{P}_{\Omega^\perp}(\underline{\Delta}) \rangle + \langle \mathcal{P}_{\Omega^\perp}(\underline{\mathbf{G}}), \mathcal{P}_{\Omega^\perp}(\underline{\Delta}) \rangle \\ &\geq -\langle \underline{\mathbf{G}}, \underline{\Delta} \rangle - \|\mathcal{P}_\Omega(\lambda \text{sign}(\underline{\mathbf{S}}_0) - \underline{\mathbf{G}})\|_{\text{F}} \|\mathcal{P}_\Omega(\underline{\Delta})\|_{\text{F}} + \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1 - \|\mathcal{P}_{\Omega^\perp}(\underline{\mathbf{G}})\|_\infty \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1 \\ &\geq -\langle \underline{\mathbf{G}}, \underline{\Delta} \rangle - \frac{\lambda}{4} \|\mathcal{P}_\Omega(\underline{\Delta})\|_{\text{F}} + (1 - \|\mathcal{P}_{\Omega^\perp}(\underline{\mathbf{G}})\|_\infty) \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1 \end{aligned} \quad (\text{A13})$$

Also note that

$$\begin{aligned} \|\mathcal{P}_\Omega(\underline{\Delta})\|_{\text{F}} &\leq \|\mathcal{P}_\Omega \mathcal{P}_\mathbf{T}(\underline{\Delta})\|_{\text{F}} + \|\mathcal{P}_\Omega \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{F}} \\ &\leq \|\mathcal{P}_\Omega \mathcal{P}_\mathbf{T}(\underline{\Delta})\|_{\text{F}} + \|\mathcal{P}_\Omega \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{F}} \\ &\leq \frac{1}{2} \|\underline{\Delta}\|_{\text{F}} + \|\mathcal{P}_\Omega \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{F}} \\ &\leq \frac{1}{2} \|\mathcal{P}_\Omega(\underline{\Delta})\|_{\text{F}} + \frac{1}{2} \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_{\text{F}} + \|\mathcal{P}_\Omega \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{F}} \end{aligned} \quad (\text{A14})$$

which leads to

$$\|\mathcal{P}_\Omega(\underline{\Delta})\|_{\text{F}} \leq \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_{\text{F}} + 2\|\mathcal{P}_\Omega \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{F}} \leq \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1 + 2\sqrt{n_3} \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{TNN}}. \quad (\text{A15})$$

Putting things together, we have

$$\begin{aligned} & \|\underline{L}_0 + \underline{\Delta}\|_{\text{TNN}} + \lambda \|\underline{S}_0 - \underline{\Delta}\|_1 - (\|\underline{L}_0\|_{\text{TNN}} + \lambda \|\underline{S}_0\|_1) \\ & \geq \left(1 - \frac{\lambda\sqrt{n_3}}{2} - \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{G})\|_{\text{sp}}\right) \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{TNN}} + \left(\frac{3}{4}\lambda - \mathcal{P}_{\Omega^\perp}(\|\underline{G}\|_\infty)\right) \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1. \end{aligned} \quad (\text{A16})$$

Since $\lambda \leq \frac{1}{2\sqrt{n_3}}$, it holds that

$$\begin{aligned} & \|\underline{L}_0 + \underline{\Delta}\|_{\text{TNN}} + \lambda \|\underline{S}_0 - \underline{\Delta}\|_1 \\ & \geq \|\underline{L}_0\|_{\text{TNN}} + \lambda \|\underline{S}_0\|_1 + \left(\frac{3}{4} - \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{G})\|_{\text{sp}}\right) \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})\|_{\text{TNN}} + \left(\frac{3}{4}\lambda - \mathcal{P}_{\Omega^\perp}(\|\underline{G}\|_\infty)\right) \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta})\|_1, \end{aligned}$$

for any perturbation $\underline{\Delta} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. $\square$

Lemma A2. Suppose that $\|\mathcal{P}_\Omega \mathcal{P}_\mathbf{T}\| \leq 1/2$, then for any $\underline{\mathbf{X}} = (\underline{X}_t, \underline{X}_s)$, we have

$$\|\mathcal{P}_\mathbf{T}(\mathcal{P}_\mathbf{T} \times \mathcal{P}_\Omega)(\underline{\mathbf{X}})\|_{\mathbb{F}, \mu}^2 \geq \frac{1 + \mu^2}{8} \|\mathcal{P}_\mathbf{T} \times \mathcal{P}_\Omega(\underline{\mathbf{X}})\|_{\mathbb{F}}^2. \quad (\text{A17})$$

Proof. According to the definitions of $\mathcal{P}_\mathbf{T}$ and $\mathcal{P}_\mathbf{T} \times \mathcal{P}_\Omega$, we have

$$\mathcal{P}_\mathbf{T}(\mathcal{P}_\mathbf{T} \times \mathcal{P}_\Omega)(\underline{\mathbf{X}}) = \left(\frac{\mathcal{P}_\mathbf{T}(\underline{X}_t) + \mathcal{P}_\Omega(\underline{X}_s)}{2}, \frac{\mathcal{P}_\mathbf{T}(\underline{X}_t) + \mathcal{P}_\Omega(\underline{X}_s)}{2} \right). \quad (\text{A18})$$

Then, we have

$$\begin{aligned} \|\mathcal{P}_\mathbf{T}(\mathcal{P}_\mathbf{T} \times \mathcal{P}_\Omega)(\underline{\mathbf{X}})\|_{\mathbb{F}, \mu}^2 &= (1 + \mu^2) \cdot \frac{1}{4} \cdot \left(\|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}}^2 + \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}}^2 + 2 \langle \mathcal{P}_\mathbf{T}(\underline{X}_t), \mathcal{P}_\Omega(\underline{X}_s) \rangle \right) \\ &= \frac{(1 + \mu^2)}{4} \left(\|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}}^2 + \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}}^2 + 2 \langle \mathcal{P}_\Omega \mathcal{P}_\mathbf{T} \mathcal{P}_\mathbf{T}(\underline{X}_t), \mathcal{P}_\Omega(\underline{X}_s) \rangle \right) \\ &\geq \frac{(1 + \mu^2)}{4} \left(\|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}}^2 + \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}}^2 - 2 \|\mathcal{P}_\Omega \mathcal{P}_\mathbf{T}\| \|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}} \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}} \right) \\ &\geq \frac{(1 + \mu^2)}{4} \left(\|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}}^2 + \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}}^2 - 2 \cdot \frac{1}{2} \|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}} \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}} \right) \\ &\geq \frac{(1 + \mu^2)}{4} \left(\|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}}^2 + \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}}^2 - \frac{\|\mathcal{P}_\mathbf{T}(\underline{X}_t)\|_{\mathbb{F}}^2 + \|\mathcal{P}_\Omega(\underline{X}_s)\|_{\mathbb{F}}^2}{2} \right) \\ &= \frac{(1 + \mu^2)}{8} \|\mathcal{P}_\mathbf{T} \times \mathcal{P}_\Omega(\underline{\mathbf{X}})\|_{\mathbb{F}}^2. \end{aligned} \quad (\text{A19})$$

Hence completes the proof. $\square$

Appendix A.1.2. Proof of Theorem 1

Proof. For $\underline{\mathbf{X}} = (\underline{L}, \underline{S})$, define $\|\underline{\mathbf{X}}\|_\diamond = \|\underline{L}\|_{\text{TNN}} + \lambda \|\underline{S}\|_1$. Let $\hat{\underline{\mathbf{X}}} = (\hat{\underline{L}}, \hat{\underline{S}})$, $\underline{\mathbf{X}}^* = (\underline{L}_0, \underline{S}_0)$. According to the optimality of $(\hat{\underline{L}}, \hat{\underline{S}})$ and the feasibility of $(\underline{L}_0, \underline{S}_0)$, we directly have

$$\|\hat{\underline{\mathbf{X}}}\|_\diamond \leq \|\underline{\mathbf{X}}^*\|_\diamond, \quad (\text{A20})$$

$$\|\hat{\underline{L}} + \hat{\underline{S}} - \underline{\mathbf{M}}\|_{\mathbb{F}} \leq \delta, \quad (\text{A21})$$

$$\|\underline{L}_0 + \underline{S}_0 - \underline{\mathbf{M}}\|_{\mathbb{F}} \leq \delta. \quad (\text{A22})$$

Let $\underline{\Delta}_t = \hat{\underline{L}} - \underline{L}_0$, $\underline{\Delta}_s = \hat{\underline{S}} - \underline{S}_0$. Then, we have

$$\|\underline{\Delta}_t + \underline{\Delta}_s\|_{\mathbb{F}} = \|\hat{\underline{L}} + \hat{\underline{S}} - \underline{\mathbf{M}} - (\underline{L}_0 + \underline{S}_0 - \underline{\mathbf{M}})\|_{\mathbb{F}} \leq \|\hat{\underline{L}} + \hat{\underline{S}} - \underline{\mathbf{M}}\|_{\mathbb{F}} + \|\underline{L}_0 + \underline{S}_0 - \underline{\mathbf{M}}\|_{\mathbb{F}} \leq 2\delta. \quad (\text{A23})$$

Define the pair of error tensors $\underline{\Delta} = \hat{\underline{X}} - \underline{X}^* = (\underline{\Delta}_l, \underline{\Delta}_s)$. The goal is to bound $\|\underline{\Delta}\|_{F,\mu}$.

First, we use the decomposition $\underline{\Delta} = \mathcal{P}_{\mathbf{T}}(\underline{\Delta}) + \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta})$, and let $\underline{\Delta}^\Gamma = \mathcal{P}_{\mathbf{T}}(\underline{\Delta}) = (\underline{\Delta}_l^\Gamma, \underline{\Delta}_s^\Gamma) = (\frac{\underline{\Delta}_l + \underline{\Delta}_s}{2}, \frac{\underline{\Delta}_l - \underline{\Delta}_s}{2})$, $\underline{\Delta}^{\Gamma^\perp} = \mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}) = (\underline{\Delta}_l^{\Gamma^\perp}, \underline{\Delta}_s^{\Gamma^\perp}) = (\frac{\underline{\Delta}_l - \underline{\Delta}_s}{2}, \frac{\underline{\Delta}_s - \underline{\Delta}_l}{2})$ for simplicity. Then, we have

$$\|\underline{\Delta}\|_{F,\mu} = \|\underline{\Delta}^\Gamma + \underline{\Delta}^{\Gamma^\perp}\|_{F,\mu} \leq \|\underline{\Delta}^\Gamma\|_{F,\mu} + \|\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu}. \quad (\text{A24})$$

Please note that $\underline{\Delta}_l^\Gamma = \underline{\Delta}_s^\Gamma = \frac{\underline{\Delta}_l + \underline{\Delta}_s}{2}$, thus $\|\underline{\Delta}^\Gamma\|_{F,\mu}$ can be bounded easily as follows

$$\|\underline{\Delta}^\Gamma\|_{F,\mu} = \sqrt{\|\underline{\Delta}_l^\Gamma\|_F^2 + \mu^2 \|\underline{\Delta}_s^\Gamma\|_F^2} = \frac{\sqrt{1+\mu^2}}{2} \|\underline{\Delta}_l + \underline{\Delta}_s\|_F \leq \delta \sqrt{1+\mu^2}. \quad (\text{A25})$$

Then, it remains to bound $\|\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu}$. Due to the triangular inequality

$$\|\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu} \leq \|(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu} + \|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu}, \quad (\text{A26})$$

(A) bound $\|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu}$. According to the convexity of $\|\cdot\|_\diamond$ we have

$$\|\underline{X}^* + \underline{\Delta}\|_\diamond = \|\underline{X}^* + \underline{\Delta}^\Gamma + \underline{\Delta}^{\Gamma^\perp}\|_\diamond \geq \|\underline{X}^* + \underline{\Delta}^\Gamma\|_\diamond - \|\underline{\Delta}^{\Gamma^\perp}\|_\diamond. \quad (\text{A27})$$

Using Lemma A1, we have

$$\begin{aligned} \|\underline{X}^* + \underline{\Delta}^{\Gamma^\perp}\|_\diamond &\geq \|\underline{X}^*\|_\diamond + \left(\frac{3}{4} - \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{G})\|_{\text{sp}}\right) \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}_l^{\Gamma^\perp})\|_{\text{TNN}} + \left(\frac{3}{4}\lambda - \mathcal{P}_{\Omega^\perp}(\|\underline{G}\|_\infty)\right) \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta}_s^{\Gamma^\perp})\|_1 \\ &\geq \|\underline{X}^*\|_\diamond + \frac{1}{4} \|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_\diamond. \end{aligned} \quad (\text{A28})$$

Combining Equations (A20), (A27) and (A28), we have

$$\|\underline{\Delta}^\Gamma\|_\diamond \geq \frac{1}{4} \|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_\diamond \quad (\text{A29})$$

Then, with $\mu = \sqrt{n_3}\lambda$, we reach a bound on $\|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu}$ as follows

$$\begin{aligned} \|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu} &\leq \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}_l^{\Gamma^\perp})\|_F + \mu \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta}_s^{\Gamma^\perp})\|_F \\ &\leq \sqrt{n_3} \|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}_l^{\Gamma^\perp})\|_{\text{TNN}} + \mu \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta}_s^{\Gamma^\perp})\|_1 \\ &\leq \sqrt{n_3} \left(\|\mathcal{P}_{\mathbf{T}^\perp}(\underline{\Delta}_l^{\Gamma^\perp})\|_{\text{TNN}} + \lambda \|\mathcal{P}_{\Omega^\perp}(\underline{\Delta}_s^{\Gamma^\perp})\|_1 \right) \\ &\leq \sqrt{n_3} \|(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})\underline{\Delta}^{\Gamma^\perp}\|_\diamond \\ &\leq 4\sqrt{n_3} \|\underline{\Delta}^\Gamma\|_\diamond \\ &\leq 4\sqrt{n_3} \left(\|\underline{\Delta}_l^\Gamma\|_{\text{TNN}} + \lambda \|\underline{\Delta}_s^\Gamma\|_1 \right) \\ &\leq 4\sqrt{n_3} \left(\sqrt{\min\{n_1, n_2\}} \|\underline{\Delta}_l^\Gamma\|_F + \lambda \sqrt{n_1 n_2 n_3} \|\underline{\Delta}_s^\Gamma\|_F \right) \\ &= 4\sqrt{n_3} \left(\sqrt{\min\{n_1, n_2\}} + \lambda \sqrt{n_1 n_2 n_3} \right) \|\underline{\Delta}_l^\Gamma\|_F \\ &\leq 4\sqrt{n_3} \left(\sqrt{\min\{n_1, n_2\}} + \lambda \sqrt{n_1 n_2 n_3} \right) \delta. \end{aligned} \quad (\text{A30})$$

(B) bound $\|(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})\underline{\Delta}^{\Gamma^\perp}\|_{F,\mu}$. Please note that

$$\mathcal{P}_{\mathbf{T}}(\underline{\Delta}^{\Gamma^\perp}) = \underline{0} = \mathcal{P}_{\mathbf{T}}(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})(\underline{\Delta}^{\Gamma^\perp}) + \mathcal{P}_{\mathbf{T}}(\mathcal{P}_{\mathbf{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(\underline{\Delta}^{\Gamma^\perp}), \quad (\text{A31})$$

which means

$$\|\mathcal{P}_{\Gamma}(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})(\underline{\Delta}^{\Gamma^{\perp}})\|_{F,\mu} = \|\mathcal{P}_{\Gamma}(\mathcal{P}_{\mathbf{T}^{\perp}} \times \mathcal{P}_{\Omega^{\perp}})(\underline{\Delta}^{\Gamma^{\perp}})\|_{F,\mu} \leq \|\mathcal{P}_{\mathbf{T}^{\perp}} \times \mathcal{P}_{\Omega^{\perp}}(\underline{\Delta}^{\Gamma^{\perp}})\|_{F,\mu}. \quad (\text{A32})$$

According to Lemma A2, we have

$$\begin{aligned} \|(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})(\underline{\Delta}^{\Gamma^{\perp}})\|_{F,\mu} &\leq \|\mathcal{P}_{\mathbf{T}}(\underline{\Delta}_t^{\Gamma^{\perp}})\|_F + \mu \|\mathcal{P}_{\Omega}(\underline{\Delta}_s^{\Gamma^{\perp}})\|_F \\ &\leq \sqrt{1+\mu^2} \sqrt{\|\mathcal{P}_{\mathbf{T}}(\underline{\Delta}_t^{\Gamma^{\perp}})\|_F^2 + \|\mathcal{P}_{\Omega}(\underline{\Delta}_s^{\Gamma^{\perp}})\|_F^2} \\ &= \sqrt{1+\mu^2} \|\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega}(\underline{\Delta}^{\Gamma^{\perp}})\|_F \\ &\leq \sqrt{1+\mu^2} \cdot \sqrt{\frac{8}{\sqrt{1+\mu^2}}} \cdot \|\mathcal{P}_{\Gamma}(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})(\underline{\Delta}^{\Gamma^{\perp}})\|_F. \end{aligned} \quad (\text{A33})$$

According to Equations (A32) and (A33), we obtain

$$\|(\mathcal{P}_{\mathbf{T}} \times \mathcal{P}_{\Omega})(\underline{\Delta}^{\Gamma^{\perp}})\|_{F,\mu} \leq 2\sqrt{2} \|(\mathcal{P}_{\mathbf{T}^{\perp}} \times \mathcal{P}_{\Omega^{\perp}})(\underline{\Delta}^{\Gamma^{\perp}})\|_{F,\mu}. \quad (\text{A34})$$

Thus, combining Equations (A24), (A25), (A30) and (A34), and setting $\mu = \sqrt{n_3}\lambda$, we obtain

$$\|\underline{\Delta}\|_{F,\mu} \leq \left(\sqrt{1+n_3\lambda^2} + 4(1+2\sqrt{2}) \left(\sqrt{\min\{n_1, n_2\}n_3} + n_3\lambda\sqrt{n_1n_2} \right) \right) \delta. \quad (\text{A35})$$

Since $\lambda = \frac{1}{\sqrt{\max\{n_1, n_2\}n_3}}$, we have

$$\|\underline{\Delta}\|_{F,\mu} \leq \left(\sqrt{1 + \frac{1}{\max\{n_1, n_2\}}} + 8(1+2\sqrt{2})\sqrt{\min\{n_1, n_2\}n_3} \right) \delta, \quad (\text{A36})$$

which indicates that

$$\begin{aligned} \|\hat{\underline{\mathbf{L}}} - \underline{\mathbf{L}}_0\|_F &\leq \left(\sqrt{1 + \frac{1}{\max\{n_1, n_2\}}} + 8(1+2\sqrt{2})\sqrt{\min\{n_1, n_2\}n_3} \right) \delta \\ \|\hat{\underline{\mathbf{S}}} - \underline{\mathbf{S}}_0\|_F &\leq \left(\sqrt{1 + \max\{n_1, n_2\}} + 8(1+2\sqrt{2})\sqrt{n_1n_2n_3} \right) \delta. \end{aligned} \quad (\text{A37})$$

Moreover, according to the analysis in [2], the conditions $\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbf{T}}\| \leq \frac{1}{2}$ and Equation (A8) in Lemma A1 hold with probability at least $1 - c_1(n_3 \max\{n_1, n_2\})^{-c_2}$, where c_1 and c_2 are positive constants.

In this way, the proof of Theorem 1 is completed. $\square$

Appendix A.2. Proof of Theorem 2

Proof. The key idea is to rewrite Problem (29) into a standard two-block ADMM problem. For notational simplicity, let

$$f(\mathbf{x}) = \frac{1}{2} \|\underline{\mathbf{L}} + \underline{\mathbf{S}} - \underline{\mathbf{Y}}\|_F^2, \quad g(\mathbf{z}) = \gamma \|\underline{\mathbf{K}}\|_{\text{TNN}} + \gamma \lambda R(\underline{\mathbf{S}}),$$

where $\mathbf{x}, \mathbf{y}, \mathbf{z}$ and $\mathbf{A}$ are defined as follows

$$\mathbf{x} = \begin{bmatrix} \text{vec}(\underline{\mathbf{L}}) \\ \text{vec}(\underline{\mathbf{S}}) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \text{vec}(\underline{\mathbf{Y}}_1) \\ \text{vec}(\underline{\mathbf{Y}}_2) \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \text{vec}(\underline{\mathbf{K}}) \\ \text{vec}(\underline{\mathbf{R}}) \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \text{diag}(\text{vec}(\underline{\mathbf{B}})) & \mathbf{0} \\ \mathbf{0} & \text{diag}(\text{vec}(\underline{\mathbf{B}})) \end{bmatrix},$$

and $\text{vec}(\cdot)$ denotes an operation of tensor vectorization (see [40]).

It can be verified that $f(\cdot)$ and $g(\cdot)$ are closed, proper convex functions. Then, Problem (29) can be re-written as follows:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{Ax} - \mathbf{z} = \mathbf{0}. \end{aligned}$$

According to the convergence analysis in [48], we have:

$$\begin{aligned} \text{objective convergence:} \quad & \lim_{t \rightarrow \infty} f(\mathbf{x}^t) + g(\mathbf{z}^t) = f^* + g^*, \\ \text{dual variable convergence:} \quad & \lim_{t \rightarrow \infty} \mathbf{y}^t = \mathbf{y}^*, \\ \text{constraint convergence:} \quad & \lim_{t \rightarrow \infty} \mathbf{Ax}^t - \mathbf{z}^t = \mathbf{0}, \end{aligned}$$

where f^*, g^* are the optimal values of $f(\mathbf{x}), g(\mathbf{z})$, respectively. Variable $\mathbf{y}^*$ is a dual optimal point defined as:

$$\mathbf{y}^* = \begin{bmatrix} \text{vec}(\mathbf{Y}_1^*) \\ \text{vec}(\mathbf{Y}_2^*) \end{bmatrix},$$

where $(\mathbf{Y}_1^*, \mathbf{Y}_2^*)$ is the dual component of a saddle point $(\mathbf{L}^*, \mathbf{S}^*, \mathbf{K}^*, \mathbf{R}^*, \mathbf{Y}_1^*, \mathbf{Y}_2^*)$ of the unaugmented Lagrangian $L(\mathbf{L}, \mathbf{S}, \mathbf{K}, \mathbf{R}, \mathbf{Y}_1, \mathbf{Y}_2)$. $\square$

Appendix A.3. Proof of Lemma 4

Proof. Let the full t-SVD of $\mathbf{X}$ be $\mathbf{X} = \mathbf{U} * \mathbf{\Lambda} * \mathbf{V}^\top$, where $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{r \times r \times n_3}$ are orthogonal tensors and $\mathbf{\Lambda} \in \mathbb{R}^{r \times r \times n_3}$ is f -diagonal. Then

$$\|\mathbf{X}\|_{\text{TN}} = \|\overline{\mathbf{U} * \mathbf{\Lambda} * \mathbf{V}^\top}\|_* = \|\overline{\mathbf{U}} \cdot \overline{\mathbf{\Lambda}} \cdot \overline{\mathbf{V}^\top}\|_* = \|\overline{\mathbf{\Lambda}}\|_*. \quad (\text{A38})$$

Then $\mathbf{Q} * \mathbf{X} = (\mathbf{Q} * \mathbf{U}) * \mathbf{\Lambda} * \mathbf{V}^\top$. Since

$$(\mathbf{Q} * \mathbf{U})^\top * (\mathbf{Q} * \mathbf{U}) = \mathbf{U}^\top * \mathbf{Q}^\top * \mathbf{Q} * \mathbf{U} = \mathbf{I}, \quad (\text{A39})$$

we obtain that

$$\begin{aligned} \|\mathbf{Q} * \mathbf{X}\|_{\text{TN}} &= \|\overline{\mathbf{Q} * \mathbf{X}}\|_* \\ &= \|\overline{(\mathbf{Q} * \mathbf{U}) * \mathbf{\Lambda} * \mathbf{V}^\top}\|_* \\ &= \|\overline{(\mathbf{Q} * \mathbf{U})} \cdot \overline{\mathbf{\Lambda}} \cdot \overline{\mathbf{V}^\top}\|_* \\ &= \|\overline{\mathbf{\Lambda}}\|_*. \end{aligned} \quad (\text{A40})$$

Thus, $\|\mathbf{Q} * \mathbf{X}\|_{\text{TN}} = \|\mathbf{X}\|_{\text{TN}}$. $\square$

Appendix A.4. Proof of Theorem 3

Proof. Please note that $(\mathbf{Q}_* * \mathbf{X}_*, \mathbf{S}_*)$ is a feasible point of Problem (28), then we have

$$\begin{aligned} & \frac{1}{2} \|\mathbf{B} \odot (\mathbf{L}^* + \mathbf{S}^* - \mathbf{M})\|_F^2 + \gamma(\|\mathbf{L}^*\|_{\text{TN}} + \lambda \|\mathbf{S}^*\|_1) \\ & \leq \frac{1}{2} \|\mathbf{B} \odot (\mathbf{Q}_* * \mathbf{X}_* + \mathbf{S}_* - \mathbf{M})\|_F^2 + \gamma(\|\mathbf{Q}_* * \mathbf{X}_*\|_{\text{TN}} + \lambda \|\mathbf{S}_*\|_1) \\ & = \frac{1}{2} \|\mathbf{B} \odot (\mathbf{Q}_* * \mathbf{X}_* + \mathbf{S}_* - \mathbf{M})\|_F^2 + \gamma(\|\mathbf{X}_*\|_{\text{TN}} + \lambda \|\mathbf{S}_*\|_1) \end{aligned} \quad (\text{A41})$$

By the assumption that $r_{\text{tubal}}(\mathbf{L}^*) \leq r$, there exists a decomposition $\mathbf{L}^* = \mathbf{Q}^* * \mathbf{X}^*$, such that $(\mathbf{Q}^*, \mathbf{X}^*, \mathbf{S}^*)$ is also a feasible point of Problem (39).

Moreover, since $(\underline{Q}_*, \underline{X}_*, \underline{S}_*)$ is a global optimal solution to Problem (39), then we have that

$$\begin{aligned} & \frac{1}{2} \|\underline{B} \odot (\underline{Q}_* * \underline{X}_* + \underline{S}_* - \underline{M})\|_F^2 + \gamma(\|\underline{X}_*\|_{\text{TNN}} + \lambda \|\underline{S}_*\|_1) \\ & \leq \frac{1}{2} \|\underline{B} \odot (\underline{Q}^* * \underline{X}^* + \underline{S}^* - \underline{M})\|_F^2 + \gamma(\|\underline{X}^*\|_{\text{TNN}} + \lambda \|\underline{S}^*\|_1). \end{aligned}$$

By $\underline{L}^* = \underline{Q}^* * \underline{X}^*$, we have

$$\|\underline{L}^*\|_{\text{TNN}} = \|\underline{Q}^* * \underline{X}^*\|_{\text{TNN}} = \|\underline{X}^*\|_{\text{TNN}}. \quad (\text{A42})$$

Thus, we deduce

$$\begin{aligned} & \frac{1}{2} \|\underline{B} \odot (\underline{Q}_* * \underline{X}_* + \underline{S}_* - \underline{M})\|_F^2 + \gamma(\|\underline{X}_*\|_{\text{TNN}} + \lambda \|\underline{S}_*\|_1) \\ & \leq \frac{1}{2} \|\underline{B} \odot (\underline{L}^* + \underline{S}^* - \underline{M})\|_F^2 + \gamma(\|\underline{L}^*\|_{\text{TNN}} + \lambda \|\underline{S}^*\|_1). \end{aligned} \quad (\text{A43})$$

According to Equations (A41) and (A43), we further have

$$\begin{aligned} & \frac{1}{2} \|\underline{B} \odot (\underline{Q}_* * \underline{X}_* + \underline{S}_* - \underline{M})\|_F^2 + \gamma(\|\underline{X}_*\|_{\text{TNN}} + \lambda \|\underline{S}_*\|_1) \\ & \leq \frac{1}{2} \|\underline{B} \odot (\underline{L}^* + \underline{S}^* - \underline{M})\|_F^2 + \gamma(\|\underline{L}^*\|_{\text{TNN}} + \lambda \|\underline{S}^*\|_1). \end{aligned} \quad (\text{A44})$$

In this way, $(\underline{Q}_* * \underline{X}_*, \underline{S}_*)$ is also the optimal solution to Problem (28). $\square$

References

1. Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 208–220.
2. 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.* **2019**, doi:10.1109/TPAMI.2019.2891760.
3. Xu, Y.; Hao, R.; Yin, W.; Su, Z. Parallel matrix factorization for low-rank tensor completion. *Inverse Probl. Imaging* **2015**, *9*, 601–624.
4. Liu, Y.; Shang, F. An Efficient Matrix Factorization Method for Tensor Completion. *IEEE Signal Process. Lett.* **2013**, *20*, 307–310.
5. Wang, A.; Wei, D.; Wang, B.; Jin, Z. Noisy Low-Tubal-Rank Tensor Completion Through Iterative Singular Tube Thresholding. *IEEE Access* **2018**, *6*, 35112–35128.
6. Tan, H.; Feng, G.; Feng, J.; Wang, W.; Zhang, Y.J.; Li, F. A tensor-based method for missing traffic data completion. *Transp. Res. Part C* **2013**, *28*, 15–27.
7. Peng, Y.; Lu, B.L. Discriminative extreme learning machine with supervised sparsity preserving for image classification. *Neurocomputing* **2017**, *261*, 242–252.
8. Cichocki, A.; Mandic, D.; De Lathauwer, L.; Zhou, G.; Zhao, Q.; Caiafa, C.; Phan, H.A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Process. Mag.* **2015**, *32*, 145–163.
9. Vaswani, N.; Bouwmans, T.; Javed, S.; Narayanamurthy, P. Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery. *IEEE Signal Process. Mag.* **2018**, *35*, 32–55.
10. Cichocki, A.; Lee, N.; Oseledets, I.; Phan, A.H.; Zhao, Q.; Mandic, D.P. Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 1 Low-Rank Tensor Decompositions. *Found. Trends® Mach. Learn.* **2016**, *9*, 249–429.
11. Yuan, M.; Zhang, C.H. On Tensor Completion via Nuclear Norm Minimization. *Found. Comput. Math.* **2016**, *16*, 1–38.
12. Candès, E.J.; Tao, T. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inf. Theory* **2010**, *56*, 2053–2080.
13. Hillar, C.J.; Lim, L. Most Tensor Problems Are NP-Hard. *J. ACM* **2009**, *60*, 45.

14. Yuan, M.; Zhang, C.H. Incoherent Tensor Norms and Their Applications in Higher Order Tensor Completion. *IEEE Trans. Inf. Theory* **2017**, *63*, 6753–6766.
15. Tomioka, R.; Suzuki, T. Convex tensor decomposition via structured Schatten norm regularization. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 1331–1339.
16. Semerci, O.; Hao, N.; Kilmer, M.E.; Miller, E.L. Tensor-Based Formulation and Nuclear Norm Regularization for Multienergy Computed Tomography. *IEEE Trans. Image Process.* **2014**, *23*, 1678–1693.
17. Mu, C.; Huang, B.; Wright, J.; Goldfarb, D. Square Deal: Lower Bounds and Improved Relaxations for Tensor Recovery. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 73–81.
18. Zhao, Q.; Meng, D.; Kong, X.; Xie, Q.; Cao, W.; Wang, Y.; Xu, Z. A Novel Sparsity Measure for Tensor Recovery. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 271–279.
19. Wei, D.; Wang, A.; Wang, B.; Feng, X. Tensor Completion Using Spectral (k, p) -Support Norm. *IEEE Access* **2018**, *6*, 11559–11572.
20. Tomioka, R.; Hayashi, K.; Kashima, H. Estimation of low-rank tensors via convex optimization. *arXiv* **2010**, arXiv:1010.0789.
21. Chretien, S.; Wei, T. Sensing tensors with Gaussian filters. *IEEE Trans. Inf. Theory* **2016**, *63*, 843–852.
22. Ghadermarzy, N.; Plan, Y.; Özgür Yılmaz. Near-optimal sample complexity for convex tensor completion. *arXiv* **2017**, arXiv:1711.04965.
23. Ghadermarzy, N.; Plan, Y.; Özgür Yılmaz. Learning tensors from partial binary measurements. *arXiv* **2018**, arXiv:1804.00108.
24. Liu, Y.; Shang, F.; Fan, W.; Cheng, J.; Cheng, H. Generalized Higher-Order Orthogonal Iteration for Tensor Decomposition and Completion. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 8–13 December 2014; pp. 1763–1771.
25. Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3842–3849.
26. 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, Stockholm, Sweden, 13–19 July 2018; pp. 1948–1954.
27. Jiang, J.Q.; Ng, M.K. Exact Tensor Completion from Sparsely Corrupted Observations via Convex Optimization. *arXiv* **2017**, arXiv:1708.00601.
28. Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; Qu, Y. On Unifying Multi-view Self-Representations for Clustering by Tensor Multi-rank Minimization. *Int. J. Comput. Vis.* **2018**, *126*, 1157–1179.
29. Ely, G.T.; Aeron, S.; Hao, N.; Kilmer, M.E. 5D seismic data completion and denoising using a novel class of tensor decompositions. *Geophysics* **2015**, *80*, V83–V95.
30. Liu, X.; Aeron, S.; Aggarwal, V.; Wang, X.; Wu, M. Adaptive Sampling of RF Fingerprints for Fine-grained Indoor Localization. *IEEE Trans. Mob. Comput.* **2016**, *15*, 2411–2423.
31. Wang, A.; Lai, Z.; Jin, Z. Noisy low-tubal-rank tensor completion. *Neurocomputing* **2019**, *330*, 267–279.
32. Sun, W.; Chen, Y.; Huang, L.; So, H.C. Tensor Completion via Generalized Tensor Tubal Rank Minimization using General Unfolding. *IEEE Signal Process. Lett.* **2018**, *25*, 868–872.
33. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 148–172.
34. Liu, X.Y.; Aeron, S.; Aggarwal, V.; Wang, X. Low-tubal-rank tensor completion using alternating minimization. *arXiv* **2016**, arXiv:1610.01690.
35. Liu, X.Y.; Wang, X. Fourth-order tensors with multidimensional discrete transforms. *arXiv* **2017**, arXiv:1705.01576.
36. Gu, Q.; Gui, H.; Han, J. Robust tensor decomposition with gross corruption. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 8–13 December 2014; pp. 1422–1430.
37. Wang, A.; Jin, Z.; Tang, G. Robust tensor decomposition via t-SVD: Near-optimal statistical guarantee and scalable algorithms. *Signal Process.* **2020**, *167*, 107319.
38. Zhang, Z.; Aeron, S. Exact Tensor Completion Using t-SVD. *IEEE Trans. Signal Process.* **2017**, *65*, 1511–1526.

39. Goldfarb, D.; Qin, Z. Robust low-rank tensor recovery: Models and algorithms. *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 225–253.
40. Kolda, T.G.; Bader, B.W. Tensor decompositions and applications. *SIAM Rev.* **2009**, *51*, 455–500.
41. Cheng, L.; Wu, Y.C.; Zhang, J.; Liu, L. Subspace identification for DOA estimation in massive/full-dimension MIMO systems: Bad data mitigation and automatic source enumeration. *IEEE Trans. Signal Process.* **2015**, *63*, 5897–5909.
42. Cheng, L.; Xing, C.; Wu, Y.C. Irregular Array Manifold Aided Channel Estimation in Massive MIMO Communications. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 974–988.
43. Zhao, Q.; Zhou, G.; Zhang, L.; Cichocki, A.; Amari, S.I. Bayesian robust tensor factorization for incomplete multiway data. *IEEE Trans. Neural Networks Learn. Syst.* **2016**, *27*, 736–748.
44. Zhou, Y.; Cheung, Y. Bayesian Low-Tubal-Rank Robust Tensor Factorization with Multi-Rank Determination. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, doi:10.1109/TPAMI.2019.2923240.
45. Zhou, Z.; Li, X.; Wright, J.; Candes, E.; Ma, Y. Stable principal component pursuit. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 12–18 June 2010; pp. 1518–1522.
46. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 11.
47. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5249–5257.
48. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122.
49. Peng, Y.; Lu, B.L. Robust structured sparse representation via half-quadratic optimization for face recognition. *Multimed. Tools Appl.* **2017**, *76*, 8859–8880.
50. Liu, G.; Yan, S. Active subspace: Toward scalable low-rank learning. *Neural Comput.* **2012**, *24*, 3371–3394.
51. Wang, A.; Jin, Z.; Yang, J. In *A Factorization Strategy for Tensor Robust PCA*; ResearchGate: Berlin, Germany, 2019.
52. Jiang, Q.; Ng, M. Robust Low-Tubal-Rank Tensor Completion via Convex Optimization. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 2649–2655.
53. Kernfeld, E.; Kilmer, M.; Aeron, S. Tensor–tensor products with invertible linear transforms. *Linear Algebra Its Appl.* **2015**, *485*, 545–570.
54. Lu, C.; Peng, X.; Wei, Y. Low-Rank Tensor Completion With a New Tensor Nuclear Norm Induced by Invertible Linear Transforms. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5996–6004.
55. Liu, X.Y.; Aeron, S.; Aggarwal, V.; Wang, X. Low-tubal-rank tensor completion using alternating minimization. In Proceedings of the SPIE Defense+ Security, Baltimore, MD, USA, 17–21 April 2016; International Society for Optics and Photonics: Bellingham, DC, USA, 2016; pp. 984809–984809.
56. Zhou, P.; Lu, C.; Lin, Z.; Zhang, C. Tensor Factorization for Low-Rank Tensor Completion. *IEEE Trans. Image Process.* **2018**, *27*, 1152–1163.
57. Martin, C.D.; Shafer, R.; Larue, B. An Order- p Tensor Factorization with Applications in Imaging. *SIAM J. Sci. Comput.* **2013**, *35*, A474–A490.
58. Wang, A.; Jin, Z. Orientation Invariant Tubal Nuclear Norms Applied to Robust Tensor Decomposition. Available online: https://www.researchgate.net/publication/329116872_Orientation_Invariant_Tubal_Nuclear_Norms_Applied_to_Robust_Tensor_Decomposition (accessed on 3 December 2019)

