SIAM J. IMAGING SCIENCES
Vol. 17, No. 1, pp. 756--783

A Scale-Invariant Relaxation in Low-Rank Tensor Recovery with an Application to Tensor Completion[t]

Huiwen Zheng[r], Yifei Lou[r], Guoliang Tian[r], and Chao Wang[r S]

Abstract. In this paper, we consider a low-rank tensor recovery problem. Based on the tensor singular value decomposition (t-SVD), we propose the ratio of the tensor nuclear norm and the tensor Frobenius norm (TNF) as a novel nonconvex surrogate of tensor's tubal rank. The rationale of the proposed model for enforcing a low-rank structure is analyzed as its theoretical properties. Specifically, we introduce a null space property (NSP) type condition, under which a low-rank tensor is a local minimum for the proposed TNF recovery model. Numerically, we consider a low-rank tensor completion problem as a specific application of tensor recovery and employ the alternating direction method of multipliers (ADMM) to secure a model solution with guaranteed subsequential convergence under mild conditions. Extensive experiments demonstrate the superiority of our proposed model over state-of-the-art methods.

Key words. tensor singular value decomposition, tensor completion, tensor tubal rank, null space property

MSC codes. 68Q25, 68R10, 68U05

DOI. 10.1137/23M1560847

1. Introduction.

Nowadays, with the development of modern technology, data is often collected in a multidimensional array, which is referred to as a tensor. Compared to vectors and matrices, tensors can better preserve the inherent structures in data and hence have appeared in a variety of applications including neuroscience [2, 32, 48], image science [28, 42, 49], signal processing [31, 33], seismic imaging [9, 34], and deep learning [52]. During acquisition and transmission, it is inevitable that the data is partially missing and/or affected by noise. Recovering an underlying tensor from its partial and corrupted entries is an ill-posed inverse problem, which requires reasonable assumptions posed as a regularization for guiding toward the desired solution. This paper focuses on a low-rank structure for tensor recovery. Specifically, we consider the low-rank tensor completion (LRTC) as its application, which aims to fill in the missing entries of a tensor by assuming it is low-rank. The LRTC problem has received

[t] Received by the editors April 10, 2023; accepted for publication (in revised form) January 2, 2024; published electronically March 29, 2024.
https://doi.org/10.1137/23M1560847
Funding: The second author was partially supported by NSF grant CAREER 1846690. The fourth author was partially supported by the Natural Science Foundation of China (12201286), the Shenzhen Science and Technology Program 20231115165836001, HKRGC grant CityU11301120, National Key R\&D Program of China (2023YFA1011400), and the Shenzhen Fundamental Research Program JCYJ20220818100602005.
[r] Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518005, Guangdong, Province, China (12031217@mail.sustech.edu.cn, tiangl@sustech.edu.cn, wangc6@sustech.edu.cn).
[r] Department of Mathematics and School of Data Sciences and Society at the University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (yflou@unc.edu).
[S] National Centre for Applied Mathematics Shenzhen, Shenzhen 518055, Guangdong, Province, China.

considerable attention, such as health data analysis [14, 15], image inpainting [25, 27, 35], background modeling [4], recommendation systems [12, 16], and video restoration [29].

Unlike matrices, there exist a number of definitions for tensor ranks. Two classic tensor ranks [22] are the CANDECOMP/PARAFAC (CP) rank and the Tucker rank. The CP rank, stemming from the CP decomposition [19], equals the smallest number of rank-one tensors that can be written as the outer product of vectors in appropriate dimensions. The Tucker rank [13] based on the Tucker decomposition [41] is a vector with each element being the rank of a matrix unfolded from a target tensor. Kilmer and Martin [21] proposed a tensor factorization analogous to the matrix's singular value decomposition, referred to as tensor singular value decomposition (t-SVD). Based on t-SVD, tensor multirank [13] and tubal rank [20] were introduced. In particular, tensor multirank is a vector composed of the ranks of all the frontal slices after the Fourier transforms along the tubal direction, while the tubal rank of a tensor is the maximal value of its multirank.

It is computationally intractable to minimize any of these tensor ranks directly. A popular approach for low rank approximation is to minimize a surrogate function (convex or nonconvex) to the rank. For example, Liu et al. [27] defined the sum of the nuclear norm (SNN) based on the Tucker decomposition. Besides, tensor nuclear norm (TNN) was proposed as an extension from a matrix nuclear norm in [30, 38, 53]. Unlike SNN that gives a lower bound of the Tucker rank [37], TNN is shown to be the tightest convex relaxation of the l_1-norm of the tensor multirank when confined within the unit ball of the tensor spectral norm [53].

A variety of nonconvex alternatives to TNN when approximating the rank were proposed [17, 23, 51, 54, 55]. In particular, Jiang et al. [17] extended the partial sum of singular values in the matrix case to the partial sum of the tubal nuclear norm (PSTNN), but its performance largely depends on the number of singular values to be included. In addition, Xu et al. [50] incorporated the Laplace function into TNN for LRTC. Yang et al. [51] proposed a nonconvex log-determinant function applied to tensors to modify the equal weights of singular values considered in TNN. Lu et al. [5] introduced a tensor logarithmic norm (TLN) as a nonconvex surrogate for low rankness and proposed an algorithm named logarithmic norm minimization and outlier projection (LNOP) for recovering low-rank tensors. Inspired by the matrix's Schatten-p norm, Kong, Xie, and Lin et al. [23] introduced tensor t-Schatten-p norm. Two special cases of $p = 1/2$ and $p = 2/3$ were discussed in [18] corresponding to tensor double nuclear norm and tensor Frobenius/nuclear hybrid norm, respectively. Wang et al. [46] presented a generalized nonconvex framework that can be solved by an iterative reweighted t-TNN (IR-t-TNN) algorithm.

All the aforementioned nonconvex surrogates of tensor rank have internal parameters subject to turning that largely affect the model performance. We propose a parameter-free regularization that uses the ratio of the tensor nuclear norm and the tensor Frobenius norm (TNF) to approximate the tensor tubal rank. The proposed TNF regularization is motivated by the ratio of the l_1-norm and the l_2-norm, as a scale-invariant surrogate to the l_0-norm for sparse signal recovery [36, 43, 44, 45]. In the sparse signal realm, the null space prop-erty (NSP) offers an essential criterion for exact sparse recovery by minimizing the l_1-norm. Inspired by some prior works on NSP for vectors [7, 10, 36], we introduce an NSP-type condition tailored for tensors, under which we prove that a low-rank tensor is a local minimizer of the proposed TNF model. Computationally we incorporate the TNF regularization in the

LRTC model, design an eficient algorithm via the alternating direction method of multiplier (ADMM) [3], and establish its subsequential convergence. We conduct extensive experiments using synthetic and real data for image and video inpainting to demonstrate that our proposed methods outperform state-of-the-art methods. Overall, the main contributions of our work are threefold.

- We propose a novel nonconvex regularization TNF as a parameter-free and scale-invariant surrogate to the tensor tubal rank for tensor recovery.
- We present theoretical properties of TNF, specifically showing that a low-rank tensor is a local minimum of a TNF-regularized recovery model under the NSP-type condition.
- We consider the completion as an application of tensor recovery, adapt ADMM to solve the TNF-based tensor completion model, and provide the convergence analysis.

The rest of the paper is organized as follows. Section 2 provides some notations and preliminaries. Section 3 details our proposed TNF relaxation of the tensor tubal rank with properties. The TNF-based low-rank tensor completion is discussed in section 4 with a numerical scheme and its convergence analysis. Extensive experiments are conducted in section 5 using synthetic and real data. Lastly, section 6 concludes this paper.

2. Notations and preliminaries.

This section provides an overview of fundamental notations and definitions that will be utilized throughout this paper. The field of natural numbers is denoted as $\mathbb{N}$, the field of real numbers is denoted as $\mathbb{R}$, while the field of complex numbers is denoted as $\mathbb{C}$. Tensors are denoted by boldface Euler script letters, e.g., $\mathcal{A}$. Matrices are denoted by boldface capital letters, e.g., $\mathbf{A}$; specifically, we use $\mathbf{I}$ to denote the identity matrix. Boldface lowercase letters, such as $\mathbf{a}$, represent vectors, while lowercase letters, such as a, denote scalars. The ith entry of a vector $\mathbf{a}$ is represented as a_i. For a matrix $\mathbf{A}$, its ith row and jth column are denoted by $\mathbf{A}_{i:}$ and $\mathbf{A}_{:j}$, respectively, and its (i, j)th entry is denoted by a_{ij} or $\mathbf{A}_{ij}$. Similarly, the (i, j, l)th entry of a third-order tensor $\mathcal{A}$ is denoted by a_{ijl} or $\mathcal{A}_{ijl}$. For a third-order tensor, we have column, row, and tube fibers, as $\mathcal{A}_{:jl}$, $\mathcal{A}_{i:l}$, and $\mathcal{A}_{ij:}$, respectively, while the horizontal, lateral, and frontal slides of a third-order tensor $\mathcal{A}$, are denoted by $\mathbf{A}_{i::}$, $\mathbf{A}_{:j:}$ and $\mathbf{A}_{::l}$, respectively. We use $\mathbf{A}^{(l)}$ and $\mathbf{A}_{::l}$ interchangeably to denote the lth frontal slice of $\mathcal{A}$. We use $\mathcal{O}$ to denote the zero tensor with all the entries' values being zero, and $\mathcal{N}(\mathcal{F})$ to denote the null space of a linear operator $\mathcal{F}$. The notation $[n]$ refers to a set of indexes ranging from 1 to n. We use $|S|$ to denote the cardinality of a set S.

The inner product of matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as $\langle \mathbf{A}, \mathbf{B} \rangle := \mathrm{Tr}(\mathbf{A}^* \mathbf{B})$, where $\mathbf{A}^*$ represents the conjugate transpose of $\mathbf{A}$ and $\mathrm{Tr}(\cdot)$ denotes the matrix trace. The l_2-norm of a vector $\mathbf{v} \in \mathbb{C}^n$ is defined by $\|\mathbf{v}\|_2 = \sqrt{\sum_i |v_i|^2}$. The nuclear norm of a matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$, where $\sigma_i(\mathbf{A})$ is the ith singular value of $\mathbf{A}$, while the matrix l_2 norm of $\mathbf{A}$ is defined as $\|\mathbf{A}\|_2 = \max_i \sigma_i(\mathbf{A})$. Regarding a third-order tensor, the inner product between two tensors $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{C}^{n_1\times n_2\times n_3}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{l=1}^{n_3} \langle \mathbf{A}^{(l)}, \mathbf{B}^{(l)} \rangle$. The complex conjugate of $\mathcal{A}$, which takes the complex conjugate of each entry of $\mathcal{A}$, is denoted as $\mathrm{conj}(\mathcal{A})$. The tensor l_1-norm of $\mathcal{A}$ is defined as $\|\mathcal{A}\|_1 = \sum_{ijl} |a_{ijl}|$, the infinity norm as $\|\mathcal{A}\|_\infty = \max_{ijl} |a_{ijl}|$, and the Frobenius norm as $\|\mathcal{A}\|_F = \sqrt{\sum_{ijl} |a_{ijl}|^2}$. We use $\overline{\mathcal{A}}$ to denote the tensor after applying the fast Fourier Transform (FFT) to the tensor $\mathcal{A}$ along the third (tubal) dimension, i.e., $\overline{\mathcal{A}} = \mathrm{fft}(\mathcal{A}, [\,], 3)$ via the MATLAB command "fft", and we can compute $\mathcal{A}$ back from $\overline{\mathcal{A}}$ via $\mathcal{A} = \mathrm{ifft}(\overline{\mathcal{A}}, [\,], 3)$. Let $\overline{\mathbf{A}} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$ be a block diagonal matrix of the tensor $\mathcal{A}$, i.e.,

$$\bar{A} := \text{bdiag}(\bar{\mathcal{A}}) = \begin{bmatrix} \bar{A}^{(1)} & & & \\ & \bar{A}^{(2)} & & \\ & & \ddots & \\ & & & \bar{A}^{(n_3)} \end{bmatrix}, \tag{2.1}$$

where $\bar{A}^{(i)}$ is the i-th frontal slice of $\bar{\mathcal{A}}$. Using the frontal slices of a tensor $\mathcal{A}$, we define the block circulant matrix of $\mathcal{A}$ as

$$\text{bcirc}(\mathcal{A}) := \begin{bmatrix} A^{(1)} & A^{(n_3)} & \cdots & A^{(2)} \\ A^{(2)} & A^{(1)} & \cdots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)} & A^{(n_3-1)} & \cdots & A^{(1)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}. \tag{2.2}$$

We define two operators:

$$\text{unfold}(\mathcal{A}) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n_3)} \end{bmatrix}, \quad \text{fold}(\text{unfold}(\mathcal{A})) = \mathcal{A}, \tag{2.3}$$

where unfold$(\cdot)$ maps $\mathcal{A}$ to a matrix of size $n_1 n_3 \times n_2$ and fold$(\cdot)$ is its inverse operator. As shown in [11], one has $\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{n_3}\langle \bar{A}, \bar{B} \rangle$ and $\|\mathcal{A}\|_F = \frac{1}{\sqrt{n_3}}\|\bar{A}\|_F$.

Definition 2.1 (t-product [21]). Let $\mathcal{A} \in \mathbb{R}^{n_1 \times l \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{l \times n_2 \times n_3}$; then the t-product $\mathcal{A} * \mathcal{B}$ is defined by

$$\mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})), \tag{2.4}$$

resulting in a tensor of size $n_1 \times n_2 \times n_3$. Note that $\mathcal{A} * \mathcal{B} = \mathcal{Z}$ if and only if $\bar{A}\bar{B} = \bar{Z}$.

Definition 2.2 (identity tensor [21]). The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor with its first frontal slice being the $n \times n$ identity matrix and other frontal slices being all zeros. It is clear that $\mathcal{A} * \mathcal{I} = \mathcal{A}$ and $\mathcal{I} * \mathcal{A} = \mathcal{A}$ given the appropriate dimensions.

Definition 2.3 (tensor conjugate transpose [21]). The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is a tensor $\mathcal{A}^*$ obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n_3.

Definition 2.4 (orthogonal tensor [21]). A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies $\mathcal{Q}^* * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}$.

Definition 2.5 (f-diagonal tensor [21]). A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.

Definition 2.6 (t-SVD [21]). Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; then the t-SVD of $\mathcal{A}$ is given by

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*, \tag{2.5}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}, \mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor.

It follows from Definition 2.1 that $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^t$ if and only if $\overline{A} = \overline{U}\,\overline{S}\,\overline{V}^t$.

Definition 2.7 (tensor average rank and tubal rank [30]). For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its tensor average rank, denoted as $\mathrm{rank}_a(\mathcal{A})$, is defined as

$$\mathrm{rank}_a(\mathcal{A}) = \frac{1}{n_3}\mathrm{rank}(\mathrm{bcirc}(\mathcal{A})) = \frac{1}{n_3}\sum_{i=1}^{n_3} \mathrm{rank}(\overline{A}^{(i)}).$$

The tensor tubal rank, denoted as $\mathrm{rank}_t(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ comes from the t-SVD of $\mathcal{A}$, i.e., $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^t$. In other words, one has

$$\mathrm{rank}_t(\mathcal{A}) = \#\{i, \mathcal{S}(i,i,:) \neq 0\}.$$

Definition 2.8 (tensor nuclear norm and tensor spectral norm [30]). Let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^t$ be the t-SVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Define $\sigma_{ij}(\mathcal{A})$ as the jth singular value of $\overline{A}^{(i)}$, or simply σ_{ij} if the context is clear. The TNN of $\mathcal{A}$ is defined as

$$\|\mathcal{A}\|_* := \frac{1}{n_3}\left\|\overline{A}\right\|_* = \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{\min(n_1,n_2)} \sigma_{ij}. \tag{2.6}$$

The tensor spectral norm is defined as

$$\|\mathcal{A}\| := \max_i \|\overline{A}^{(i)}\|_2 = \max_{ij} \sigma_{ij}. \tag{2.7}$$

3. Rationales of the TNF model. Despite being widely used in various low-rank tensor applications, TNN has a limitation of treating the singular values equally with the same priority, which severely suppresses the large singular values. To mitigate this drawback, we explore a nonconvex surrogate developed in sparse signal recovery by extending the ℓ_1/ℓ_2 model [36, 45] in a vector form to deal with tensors. In particular, we propose the ratio of the tensor nuclear norm and TNF, defined by

$$\|\mathcal{A}\|_{TNF} := \frac{\|\mathcal{A}\|_*}{\|\mathcal{A}\|_F}, \tag{3.1}$$

to approximate the tensor tubal rank. If we stack all the singular values $\{\sigma_{ij}\}$ as a vector, which is referred to as a singular vector, TNN is equivalent to the ℓ_1 norm of this vector up to a constant $\frac{1}{n_3}$ following (2.6). On the other hand, the Frobenius norm of a tensor is the scaled ℓ_2 norm of its singular vector, i.e.,

$$\|\mathcal{A}\|_F^2 = \frac{1}{n_3}\|\overline{A}\|_F^2 = \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{\min(n_1,n_2)} \sigma_{ij}^2. \tag{3.2}$$

As a result, TNF is analogous to the ℓ_1/ℓ_2 model [36, 45] applied on the singular vector to promote its sparsity, which effectively enforces a low-rank structure of the underlying tensor. Here we consider the tensor TNF formulation in a general tensor recovery problem given by

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{TNF} \quad \text{s.t.} \quad \mathcal{F}(\mathcal{X}) = \mathcal{T}, \tag{3.3}$$

where $\mathcal{F}(\cdot)$ is a linear operator and $\mathcal{T}$ is the corresponding measurement. We further assume that $\mathcal{T} \neq \mathcal{O}$ to avoid the trivial solution $\mathcal{X} = \mathcal{O}$. In the following, we present some theoretical properties of the proposed TNF regularization (3.1). The TNF regularization for tensors has properties inherent from ℓ_1/ℓ_2 for vectors. Here we discuss its invariance of scaling and unitary transformation, its boundedness, as well as a local optimality under an NSP-type of condition.

Proposition 3.1 (scale invariance). Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; we have

$$\|\mathcal{A}\|_{TNF} = \|c\mathcal{A}\|_{TNF},$$

which holds for any nonzero scalar c.

Proof. Without loss of generality, we assume $n_1 \leq n_2$ and denote

$$\mathbf{a} = [a_{11}, a_{12}, \ldots, a_{1n_1}, a_{21}, \ldots, a_{n_3 n_1}]^T \in \mathbb{R}^{n_1 n_3},$$

where a_{ij} is the jth singular value of $\bar{A}^{(i)}$. Based on (2.6) and (3.2), we obtain

$$\|\mathcal{A}\|_{TNF} = \sqrt{\frac{\|\mathbf{a}\|_1}{n_3}\|\mathbf{a}\|_2}. \tag{3.4}$$

For $c \neq 0$, singular values of $\overline{cA}^{(i)}$ become ca_{ij}. Owing to the scale-invariant property of ℓ_1/ℓ_2, we have

$$\|c\mathcal{A}\|_{TNF} = \sqrt{\frac{\|c\mathbf{a}\|_1}{n_3}\|c\mathbf{a}\|_2} = \sqrt{\frac{\|\mathbf{a}\|_1}{n_3}\|\mathbf{a}\|_2} = \|\mathcal{A}\|_{TNF}. \tag{3.5}$$

■

Similarly, we obtain the unitary invariance by transforming the TNF into the ratio of vector norms in terms of $\mathbf{a}$; see Proposition 3.2.

Proposition 3.2 (unitary invariance). Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; the proposed TNF regularization (3.1) satisfies

$$\|\mathcal{A}\|_{TNF} = \|\mathcal{P} * \mathcal{A}\|_{TNF} = \|\mathcal{A} * \mathcal{Q}^{*}\|_{TNF} = \|\mathcal{P} * \mathcal{A} * \mathcal{Q}^{*}\|_{TNF}$$

for any orthogonal tensors $\mathcal{P}, \mathcal{Q} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$.

It follows from Definition 2.7 that $\text{rank}_a(\mathcal{A}) = \|\mathbf{a}\|_0/n_3$. Together with $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{\|\mathbf{x}\|_0}\|\mathbf{x}\|_2$ for any vector $\mathbf{x}$, we obtain Proposition 3.3.

Proposition 3.3 (boundedness). For any nonzero tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have

$$\frac{1}{\sqrt{n_3}} \le \|\mathcal{A}\|_{TNF} \le \sqrt{\mathrm{rank}_a(\mathcal{A})} \le \min\{\sqrt{n_1}, \sqrt{n_2}\}.$$

Note that TNN can be expressed as the convex envelope of the tensor average rank within the unit ball of the tensor spectral norm [30]. Proposition 3.3 implies that the square of TNF is a nonconvex envelope of the tensor average rank for the whole domain (not limited to the unit ball of the tensor spectral norm).

In the literature of sparse vector recovery, a matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy an NSP [7, 8, 10] of order r with a constant $\rho \in (0,1)$ if the inequality

$$\|v_S\|_1 \le \rho \|v_{S^c}\|_1, \tag{3.6}$$

holds for any vector $v \in N(A)$ and any index set $S \subset [n]$ with $|S| \le r$. Tran and Webster [40] extended NSP to analyze a family of nonconvex surrogate functions, while a robust tensor NSP for tensors has been discussed in [46]. In this paper, we aim to extend the NSP analysis from the l_1 minimization to the tensor TNF formulation in a general tensor recovery problem (3.3).

Given a tensor $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and an integer $R \le n$ with $n := \min\{n_1, n_2\}$, we define two related tensors,

$$\mathcal{X}_R := \mathcal{U} * \mathcal{S}_R * \mathcal{V}^T \text{ and } \mathcal{X}_{R^c} := \mathcal{U} * \mathcal{S}_{R^c} * \mathcal{V}^T, \tag{3.7}$$

where $\mathcal{S}_R$ is a zero tensor except $\mathcal{S}_R(i,i,:) = \mathcal{S}(i,i,:)$ for $i \le R$ and $\mathcal{S}_{R^c}$ is a zero tensor except $\mathcal{S}_{R^c}(i,i,:) = \mathcal{S}(i,i,:)$ for $i \ge R+1$. After introducing a tensor version of NSP in Definition 3.4, we present a theoretical property of the proposed TNF regularization in Theorem 3.5.

Definition 3.4. We say a linear operator $\mathcal{F}$ satisfies an NSP-type condition of order r with a constant $\gamma \in (0,1)$ if the inequality

$$\|\mathcal{W}_r\|_* \le \gamma \|\mathcal{W}_{r^c}\|_* \tag{3.8}$$

holds for any tensor $\mathcal{W} \in N(\mathcal{F}) \setminus \{\mathcal{O}\}$.

Note that Definition 3.4 implies $\|\mathcal{W}_Q\|_* \le \gamma \|\mathcal{W}_{Q^c}\|_*$ for any integer $Q \le r$.

Theorem 3.5. If the linear operator $\mathcal{F}$ satisfies the NSP-type condition of order r with a positive constant $\gamma < \frac{\sqrt{r}}{\sqrt{r+n}\sqrt{n_3}}$ (Definition 3.4), then any tensor with tubal rank r and satisfying $\mathcal{F}(\mathcal{X}) = \mathcal{T}$ is a local minimizer of (3.3). In other words, there exists a constant $t_t > 0$ such that

$$\|\mathcal{X}\|_{TNF} \le \|\mathcal{X} + \mathcal{W}\|_{TNF}$$

holds for every $\mathcal{W} \in N(\mathcal{F}) \setminus \{\mathcal{O}\}$ with $\|\mathcal{W}\|_F \le t_t$.

It is straightforward to derive from the condition $\gamma < \frac{\sqrt{r}}{\sqrt{r+n}\sqrt{n_3}}$ that $\gamma < 1$, which suggests that our NSP condition (3.8) resembles the stable NSP (3.6) for vectors. The stable NSP (3.6) guarantees a stable recovery of a sparse vector from the l_1 minimization [10]. Here we show in Theorem 3.5 that any solution of $\mathcal{F}(\mathcal{X}) = \mathcal{T}$ with tubal rank upper bounded by r is a local minimizer for the TNF model (3.3) if $\mathcal{F}$ satisfies NSP in Definition 3.4. We point out that our analysis on NSP extends beyond a trivial progression from l_1/l_2 [36] to TNF, as the tensor nuclear norm is not separable through tensor addition. The proof of Theorem 3.5 is provided in Appendix A.

Remark 3.6. Note that the NSP-type condition is usually hard to be satisfied. Pfetsch and Tillmann [39] proved that it is NP-hard to verify the NSP in the matrix case [7]. Specifically for a given matrix A and positive integer r, it is NP-hard to compute the optimal constant s such that (3.8) holds for matrices.

4. TNF-based low-rank tensor completion.

Tensor completion refers to the problem of filling in missing entries of a tensor. Given the observed tensor $\mathcal{M} \in \mathbb{R}^{n_1\times n_2\times n_3}$ on the index set Ω with unknown entries taking the value of zero, the proposed LRTC model is formulated as

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{TNF} \quad \text{s.t.} \quad P_\Omega(\mathcal{X} - \mathcal{M}) = \mathcal{O}, \tag{4.1}$$

where $P_\Omega(\cdot)$ is a projection operator in a way that $P_\Omega(\mathcal{X} - \mathcal{M}) = \mathcal{O}$ forces the entries of $\mathcal{X}$ agree with $\mathcal{M}$ on Ω and sets the other elements of $\mathcal{X}$ to zero. We define the indicator function

$$I_\Gamma(\mathcal{E}) := \begin{cases} 0 & \text{if } \mathcal{E} \in \Gamma, \\ \infty & \text{otherwise,} \end{cases} \tag{4.2}$$

where $\Gamma := \{\mathcal{E} \in \mathbb{R}^{n_1\times n_2\times n_3}, P_\Omega(\mathcal{E} - \mathcal{M}) = \mathcal{O}\}$. Then (4.1) can be expressed as

$$\min_{\mathcal{X}} \frac{\|\mathcal{X}\|_*}{\|\mathcal{X}\|_F} + I_\Gamma(\mathcal{X}). \tag{4.3}$$

We apply the ADMM [3] to minimize (4.3) by rewriting it into an equivalent form

$$\min_{\mathcal{X}} \frac{\|\mathcal{X}\|_*}{\|\mathcal{H}\|_F} + I_\Gamma(\mathcal{X}) \quad \text{s.t.} \quad \mathcal{X} = \mathcal{H}, \tag{4.4}$$

with an auxiliary tensor $\mathcal{H}$ of the same dimension to $\mathcal{X}$. The augmented Lagrangian for (4.4) is given by

$$L_{TC}(\mathcal{X}, \mathcal{H}, \mathcal{A}) = \frac{\|\mathcal{X}\|_*}{\|\mathcal{H}\|_F} + I_\Gamma(\mathcal{X}) + \frac{\rho_1}{2}\|\mathcal{X} - \mathcal{H}\|_F^2 + \langle \mathcal{A}, \mathcal{X} - \mathcal{H}\rangle, \tag{4.5}$$

where $\mathcal{A}$ is a Lagrange multiplier and ρ_1 is a positive parameter. The ADMM scheme of (4.5) updates the tensors sequentially as follows:

$$\begin{cases} \mathcal{X}^{(k+1)} = \arg\min_{\mathcal{X}} L_{TC}\left(\mathcal{X}, \mathcal{H}^{(k)}, \mathcal{A}^{(k)}\right), \\ \mathcal{H}^{(k+1)} = \arg\min_{\mathcal{H}} L_{TC}\left(\mathcal{X}^{(k+1)}, \mathcal{H}, \mathcal{A}^{(k)}\right), \\ \mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} + \rho_1\left(\mathcal{X}^{(k+1)} - \mathcal{H}^{(k+1)}\right). \end{cases} \tag{4.6}$$

The X -subproblem in (4.6) is equivalent to

$$\min_{X} \frac{|X|_t}{|H^{(k)}|_F} + I_i(X) + \frac{u_1}{2}\left|X - H^{(k)} + \frac{1}{\iota_1}A^{(k)}\right|_F^2, \tag{4.7}$$

which is a TNN-regularized tensor completion problem. Hence we can apply ADMM by introducing one more auxiliary variable Y such that the X -subproblem (4.7) becomes

$$\min_{X} \frac{|X|_t}{|H^{(k)}|_F} + I_i(Y) + \frac{u_1}{2}\left|X - H + \frac{1}{\iota_1}A^{(k)}\right|_F^2 \quad \text{s.t.} \quad X = Y. \tag{4.8}$$

The augmented Lagrangian for (4.8) is given by

$$L^{(k)}(X, Y, B) = \frac{|X|_t}{|H^{(k)}|_F} + I_i(Y) + \frac{u_1}{2}\left| - H^{(k)} + \frac{1}{u_1}A^{(k)}\right|_F^2 + \frac{\iota_2}{2}\left|X - Y + \frac{1}{\iota_2}B\right|_F^2,$$

X

where B is a Lagrange multiplier and u_2 is a positive parameter. The ADMM framework leads the following iterations:

$$\begin{cases} X_{j+1} = \arg\min_X L^{(k)}(X, Y_j, B_j), \\ Y_{j+1} = \arg\min_Y L^{(k)}(X_{j+1}, Y, B \end{cases} \tag{4.9}$$

$_j)$,

$$B_{j+1} = B_j + u_2(X_{j+1} - Y_{j+1}),$$

where the subscript j indicates the inner loop index, as opposed to the superscript k for the outer iteration in (4.6).

The subproblem for updating X_{j+1} can be written as

(4.10)

$$X_{j+1} = \arg\min_X \frac{|X|_t}{(u_1 + u_2)|H^{(k)}|_F} + \frac{1}{2}\left|X - \frac{1}{u_1 + u_2}\left(u_1 H^{(k)} + u_2 Y_j\right) + \frac{A^{(k)} + E_j}{u_1 + u_2}\right|_F^2,$$

which has a closed-form solution via the tensor singular value thresholding (t-SVT) [30]. To make our paper self-contained, we present a theorem in [30], followed by the formula to update X_{j+1}.

Theorem 4.1 (t-SVT [30]). Given a third-order tensor Z n $R^{n_1 s\, n_2 s\, n_3}$ and a positive scalar u, a minimizer to the problem

$$\min_Y u\,|Y|_t + \frac{1}{2}|Y - Z|_F^2 \tag{4.11}$$

is given by $Y = D_u(Z)$ with the t-SVT operator $D_u(t)$ defined by

$$D_u(Z) := U \, t \, S_u \, t \, V^t,$$

where $Z = U\, t\, S\, t\, V^t$ and S_u is an $n_1 s\; n_2 s\; n_3$ tensor that satisfies $\overline{S}_u = \max\{\overline{S} - u, 0\}$.

Comparing (4.10) and (4.11), we have the formula

$$X_{j+1} = D_u(Z_j), \tag{4.12}$$

where $u = \frac{1}{(u_1 + u_2)\|H^{(k)}\|_F}$ and $Z_j = \frac{1}{u_1+u_2}\left(u_1 H^{(k)} + u_2 Y_j\right) - \frac{A^{(k)} + B_j}{u_1 + u_2}$.

The Y-subproblem in (4.9) is expressed by

$$Y_{j+1} = \arg\min_{Y} \left\{ l_i(Y) + \frac{u_2}{2}\left\| Y - \left(X_{j+1} + \frac{1}{u_2} B_j \right) \right\|_F^2 \right\}, \tag{4.13}$$

which amounts to projecting the tensor $X_{j+1} + \frac{1}{u_2} B_j$ onto the set a^C, i.e.,

$$\begin{cases} P_a(Y_{j+1}) = P_a(M), \\ P_{a^c}(Y_{j+1}) = P_{a^c}\left(X_{j+1} + \frac{1}{u_2} B_j \right), \end{cases} \tag{4.14}$$

where a^C denotes the complementary set of a. Note that (4.14) can be rewritten as

$$Y_{j+1} = P_a(M) + P_{a^c}\left(X_{j+1} + \frac{1}{u_2} B_j \right).$$

Now, we turn to the H-subproblem in the outer loop (4.6), which can be written as

$$H^{(k+1)} = \arg\min_{H} \frac{\|X^{(k+1)}\|_t}{\|H\|_F} + \frac{u_1}{2}\left\| H - \left(X^{(k+1)} + \frac{A^{(k)}}{u_1} \right) \right\|_F^2. \tag{4.15}$$

In order to solve (4.15), we consider the following optimization problem:

$$H^{(k+1)} = \arg\min_{H} \left\{ \frac{o^{(k+1)}}{\|H\|_F} + \frac{\iota_1}{2}\left\| H - K^{(k)} \right\|_F^2 \right\}, \tag{4.16}$$

with a scalar $o^{(k+1)} = \|X^{(k+1)}\|_t$ and a tensor $K^{(k)} = X^{(k+1)} + \frac{A^{(k)}}{\iota_1}$. Following the work of [36], we derive the closed-form solution to the problem (4.16) given by

$$H^{(k+1)} = \begin{cases} a^{(k)} K^{(k)} & \text{if } K^{(k)} \neq O, \\ G^{(k)} & \text{otherwise}, \end{cases} \tag{4.17}$$

where $G^{(k)}$ is a random tensor with its Frobenius norm being $\sqrt[3]{\frac{o^{(k+1)}}{\iota_1}}$, and $a^{(k)} = \frac{1}{3} + \frac{1}{3}\left(C^{(k)} + \frac{1}{C^{(k)}}\right)$ with

$$C^{(k)} = \sqrt[3]{\frac{27E^{(k)} + 2 + \sqrt{(27E^{(k)} + 2)^2 - 4}}{2}} \quad \text{and} \quad E^{(k)} = \frac{o^{(k+1)}}{u_1 \|K^{(k)}\|_F^3}.$$

We summarize the ADMM scheme in Algorithm 4.1 for solving the TNF-based tensor completion problem (4.1).

Algorithm r.e. The TNF-based tensor completion via ADMM.

Require: Observed data M on the index set a, parameters: u_1, u_2, kMax, jMax, n.

```
Initialization: choose the TNN-regularized LRTC solution to get X^(0), set H^(0) = X^(0)
  and k = 1.
while k q kMax or not converged do
  while j q jMax do
    X_{j+1} = D_{u^(k)/(u_1+u_2)} ( (u_1 H^(k) + u_2 Y_j)/(u_1+u_2) - (A^(k) + B_j)/(u_1+u_2) )
    Y_{j+1} = P_a(M) + P_{a^c}(X_{j+1} + (1/u_2) B_j)
    B_{j+1} = B_j + u_2 (X_{j+1} - Y_{j+1})
    j = j + 1
  end while
  return X^(k+1) = X_{j+1}, Y^(k+1) = Y_{j+1}
  H^(k+1) = { a^(k) (X^(k+1) + A^(k)/u_1)   if X^(k+1) + A^(k)/u_1 = O
            { G^(k)                          otherwise
  A^(k+1) = A^(k) + u_1 (X^(k+1) - H^(k+1))
  u^(k+1) = 1 / ((u_1 + u_2) |H^(k+1)|_F)
  k = k + 1
  Check the convergence conditions
    |X^(k+1) - X^(k)|_y q n,  |A^(k+1) - A^(k)|_y q n
end while
return X^ = X^(k)
```

4.1. Complexity. Here we discuss the computational complexity of Algorithm 4.1, where the computations are dominated by updating X and H. For the X-update, there is an inner loop in every iteration, in which the computation cost is mainly from t-SVT with complexity of $O(n_1 n_2 n_3 (\log n_3 + \min\{n_1, n_2\}))$. Therefore, the total computational cost of updating X is $O((\text{jMax } t\ n_1 n_2 n_3 (\log n_3 + \min\{n_1, n_2\}))$, where jMax is the iteration number of the inner loop. To update H, it takes $O(n_1 n_2 n_3 \min\{n_1, n_2\})$ for t-SVD. All together, the total cost per outer iteration in Algorithm 4.1 is at most

$$O(\text{jMax } t\ n_1 n_2 n_3 \log n_3 + (\text{jMax} + 1) n_1 n_2 n_3 \min\{n_1, n_2\}).$$

4.2. Convergence analysis. This section is devoted to the convergence analysis of Algorithm 4.1 for LRTC. In particular, we show that the sequence generated by (4.6) has a subsequence convergent to a stationary point under the following two assumptions.

C1: The sequence $\{X^{(k)}\}$ generated by (4.6) is bounded, and hence its nuclear norm is also bounded, which is denoted as $\sup_k \{|X^{(k)}|_t\}$ q M.

C2: The Frobenius norm of $\{H^{(k)}\}$ has a uniform lower bound, i.e., there exists a constant $a > 0$ such that $|H^{(k)}|_F$ q a l k.

Lemma 4.2. Under assumptions C1--C2, the sequence $\{X^{(k)}, H^{(k)}, A^{(k)}\}$ generated by (4.6) satisfies

$$\left\|A^{(k+1)} - A^{(k)}\right\|_F^2 \leq \frac{2n}{a^4}\left\|X^{(k+1)} - X^{(k)}\right\|_F^2 + \frac{4M^2}{a^6}\left\|H^{(k+1)} - H^{(k)}\right\|_F^2, \tag{4.18}$$

where $n := \min\{n_1, n_2\}$, and M and a are the constants defined in C1 and C2, respectively.

Lemma 4.3 (suficient descent). Under assumptions C1--C2 and a suficiently large param-eter $u_1 > 0$, the augmented Lagrangian function of the sequence $\{X^{(k)}, H^{(k)}, A^{(k)}\}$ generated by (4.5) satisfies

$$\begin{aligned} L_{T}C}\left(X^{(k+1)}, H^{(k+1)}, A^{(k+1)}\right) \leq\ & L_{T}C}\left(X^{(k)}, H^{(k)}, A^{(k)}\right) \\ & - c_1\left\|X^{(k+1)} - X^{(k)}\right\|_F^2 - c_2\left\|H^{(k+1)} - H^{(k)}\right\|_F^2, \end{aligned} \tag{4.19}$$

where c_1, c_2 are two positive constants.

Lemma 4.4. Let $\{X^{(k)}, H^{(k)}, A^{(k)}\}$ be the sequence generated by (4.6); then there exists a tensor $W^{(k+1)} \in \partial L_{T}C}(X^{(k+1)}, H^{(k+1)}, A^{(k+1)})$ and two constants $a_1, a_2 > 0$ such that

$$\left\|W^{(k+1)}\right\|_F^2 \leq a_1\left\|X^{(k+1)} - X^{(k)}\right\|_F^2 + a_2\left\|H^{(k+1)} - H^{(k)}\right\|_F^2. \tag{4.20}$$

Theorem 4.5. Under assumptions C1--C2 and a suficiently large parameter $u_1 > 0$, the sequence $\{X^{(k)}, H^{(k)}, A^{(k)}\}$ generated by (4.6) satisfies the following:

(i) The sequences $\{H^{(k)}\}$ and $\{A^{(k)}\}$ are bounded.

(ii) $\|X^{(k+1)} - X^{(k)}\|_F \to 0$, $\|H^{(k+1)} - H^{(k)}\|_F \to 0$, $\|A^{(k+1)} - A^{(k)}\|_F \to 0$, as $k \to \infty$.

(iii) $\{X^{(k)}, H^{(k)}, A^{(k)}\}$ has a subsequence convergent to a critical point $\{X^t, H^t, A^t\}$, namely $O \in \partial L_{T}C}(X^t, H^t, A^t)$, where O is a zero tensor but indeed composed of three zero ten-sors, each of dimension $n_1 \times n_2 \times n_3$.

Remark 4.6. Since the proposed TNF regularization (3.1) is not coercive, the boundedness assumption (C1) is required for the convergence proof. In fact, it is possible that the optimal solution could be unbounded if the sampling set a is not the entire domain. In addition, the uniform lower bound a in the assumption C2 is important in (4.18) to guarantee the suficient descent as in Lemma 4.3. Both assumptions (C1 and C2) seem strong from the theoretical point of view, but they can be verified numerically.

Theorem 4.5 characterizes the subsequential convergence. We further show that the aug-mented Lagrangian $L_{T}C}$ has the Kurdyka-\Lojasiewicz (KL) property so that the global conver-gence can be established in Theorem 4.9. To make our paper self-contained, the KL property is defined as follows.

Definition 4.7 (KL property [1]). We say a proper closed function $h : \mathbb{R}^n \to (-\infty, +\infty]$ sat-isfies the KL property at a point $\hat{x} \in \operatorname{dom} \partial h$ if there exist a constant $u \in (0, \infty]$, a neighborhood U of $\hat{x}$, and a continuous concave function $i : [0, u) \to [0, \infty)$ with $i(0) = 0$ such that

(a) i is continuously differentiable on $(0, u)$ with $i' > 0$ on $(0, u)$;

(b) for every $x \in U$ with $h(\hat{x}) < h(x) < h(\hat{x}) + u$, it holds that

$$i'(h(x) - h(\hat{x}))\,\operatorname{dist}(o, \partial h(x)) \leq 1,$$

where $\mathrm{dist}(x, C)$ denotes the distance from a point x to a closed set C measured in $\| \cdot \|_2$ with a convention of $\mathrm{dist}(x, \emptyset) := +\infty$.

To establish the global convergence of ADMM, we consider a modified augmented Lagrangian, denoted by L_a, by incorporating the lower bound of $\|H\|_F$, i.e.,

$$L_a(X, H, A) = \frac{\|X\|_*}{\|H\|_F} + I_{\Omega}(X) + \frac{u_1}{2}\|X - H\|_F^2 + \langle A, X - H\rangle + I_{\|H\|_F \geq a}(H). \quad (4.21)$$

Note that $L_a = L_{\beta}$ under the assumption C2. Now we further show that L_a has the KL property, which implies that L_{β} satisfies the KL property under C2.

Lemma 4.8. L_a defined in (4.5) satisfies the KL property.

Theorem 4.9 (global convergence). Under assumptions C1--C2 and a suficiently large $u_1 > 0$, the sequence $\{X^{(k)}, H^{(k)}, A^{(k)}\}$ generated by (4.6) converges to a stationary point of (4.5).

The proofs of Lemmas 4.2, 4.3, and 4.4 and Theorem 4.5 are provided in Appendix B. As the proofs of Lemma 4.8 and Theorem 4.9 follow almost the same as [43, Lemma 3] and [24, Theorem 4], respectively, we omit them here.

5. Experiments.

In this section, we conduct extensive experiments to evaluate the performance of the proposed TNF regularization using both synthetic and real-world data, showing its superiority over state-of-the-art methods for the LRTC problem. All the experiments are implemented using MATLAB (R2022b) on the Windows 10 platform with Intel Core i5-1135G7 2.40 GHz and 16 GB of RAM.

5.1. Synthetic data.

We generate a ground-truth low rank tensor by t-product, i.e., $X_{GT} = P * Q$, where $P \in \mathbb{R}^{n\times r\times n_3}$ and $Q \in \mathbb{R}^{r\times n\times n_3}$ with $r \ll n$, and the tubal rank of the resulting tensor $X_{GT} \in \mathbb{R}^{n\times n\times n_3}$ is at most r. The elements of tensors P and Q are drawn from an i.i.d. Gaussian distribution $N(0, \frac{1}{n})$. We compare the proposed TNF regularization with TNN [53], PSTNN [17], Laplace function based nonconvex surrogate [50] (labelled by ``Laplace''), LNOP [5], and IR-t-TNN [46]. For the TNF regularized model (4.1), we set n as 10^{-4} for LRTC in Algorithm 4.1. In addition, we follow the work of [30] to gradually increase the values of u_1 and u_2, rather than fixing their values for acceleration. For the competing methods, we use the MATLAB codes provided by respective authors with default parameter settings.

We consider the success rates to evaluate the recovery performance, which is defined by the ratio of successful trials over the total number of trials. Specifically, for every pair of a preset tubal rank and a sampling ratio, we generate ten independent random trials. For a trial to be considered successful, the relative square error between the recovered tensor $\hat{X}$ and the ground-truth tensor X_{GT}, i.e., $\|\hat{X} - X_{GT}\|_F^2/\|X_{GT}\|_F^2$, is less than 10^{-3}. The success rate is then calculated by dividing the number of successful trials by 10.

We start by a discussion about the influence of jMax in Algorithm 4.1 on the performance of low-rank tensor completion. For this purpose, we adopt a third-order tensor of dimension $40\times 40\times 20$, tubal rank 9, and sampling rate 0.5. The maximum outer iterations are set to 500, while we analyze the outcomes for jMax values of 1, 3, 5, and 10. In Figure 1, we

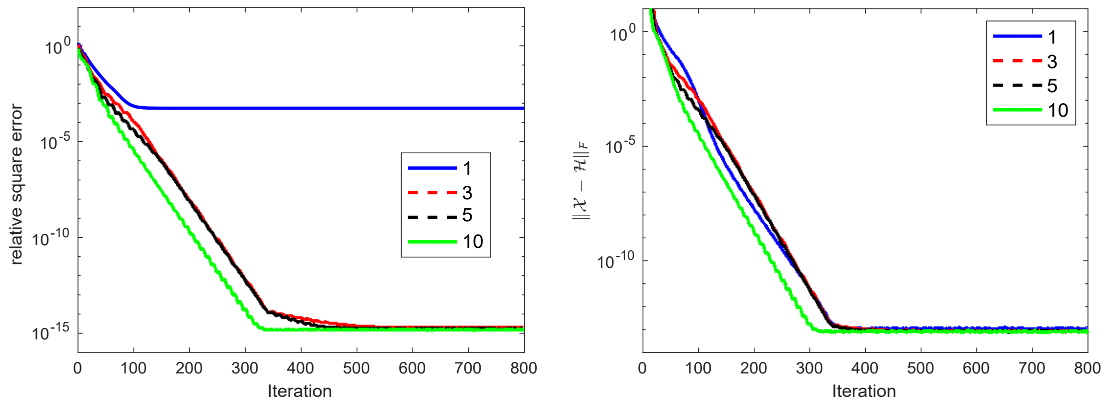

Figure 1. The influence of the maximum number of the inner loops on the relative square error (left) between the current tensor X and the ground truth X G}T} and the difference (right) between X and its counterpart H with respect to every iteration (counting both inner and outer loops).

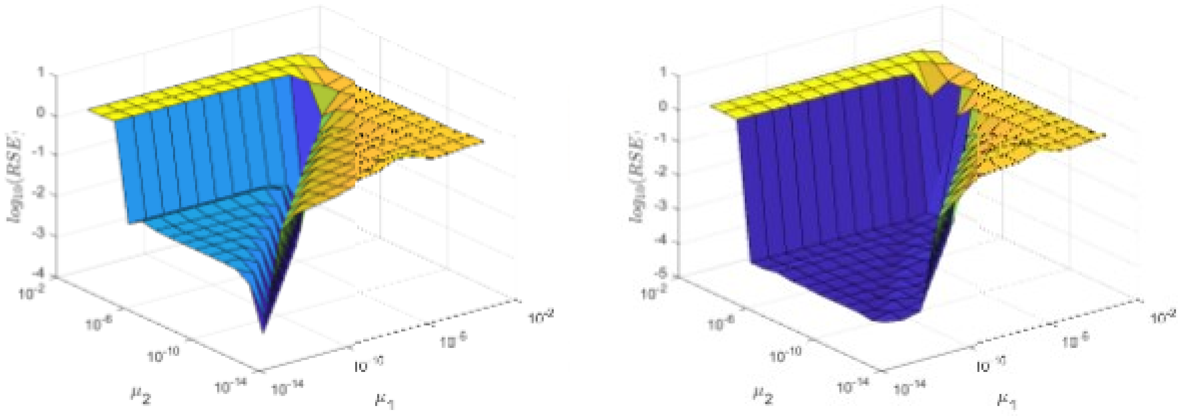

Figure 2. Sensitivity analysis on the parameters u_1 and u_2 in Algorithm 4.1. The ground-truth tensor is of tubal rank 9. We plot RSE in a logarithmic scale with sampling rates of 0.5 (left) and 0.6 (right), showing that our algorithm is not sensitive to u_1,u_2 if chosen within a proper range.

plot the relative square error (RSE) of the current tensor X to the ground truth X G}T} and the difference to its auxiliary counterpart H across iterations (counting both inner and outer loops). When jMax = 1, the relative square error does not decrease suficiently compared to the other cases. For jMax q 3, all the curves coincide after 500 iterations. Since more inner iterations require longer computational time, we set jMax as 3 for the rest of the experiments.

Subsequently, we examine the sensitivity of the parameters u_1 and u_2 in Algorithm 4.1. In a similar vein, we utilize two third-order tensors, both of dimension 40 s 40 s 20 and tubal rank of 9. We choose two sampling ratios of 0.5 and 0.6. For the parameter exploration, we consider variations within the range (u_1,u_2) n (10_i, 10_j), where i and j range from - 14 to - 2. The resulting RSE is shown in Figure 2. Notably, our observations indicate favorable parameter performance when u_1,u_2 q 10^{-3} and u_1 q u_2. In addition, the best performance emerges when the values of u_1 and u_2 are in the similar range. Moreover, the sensitivity of our algorithm on parameters reduces further as the sampling rate increases.

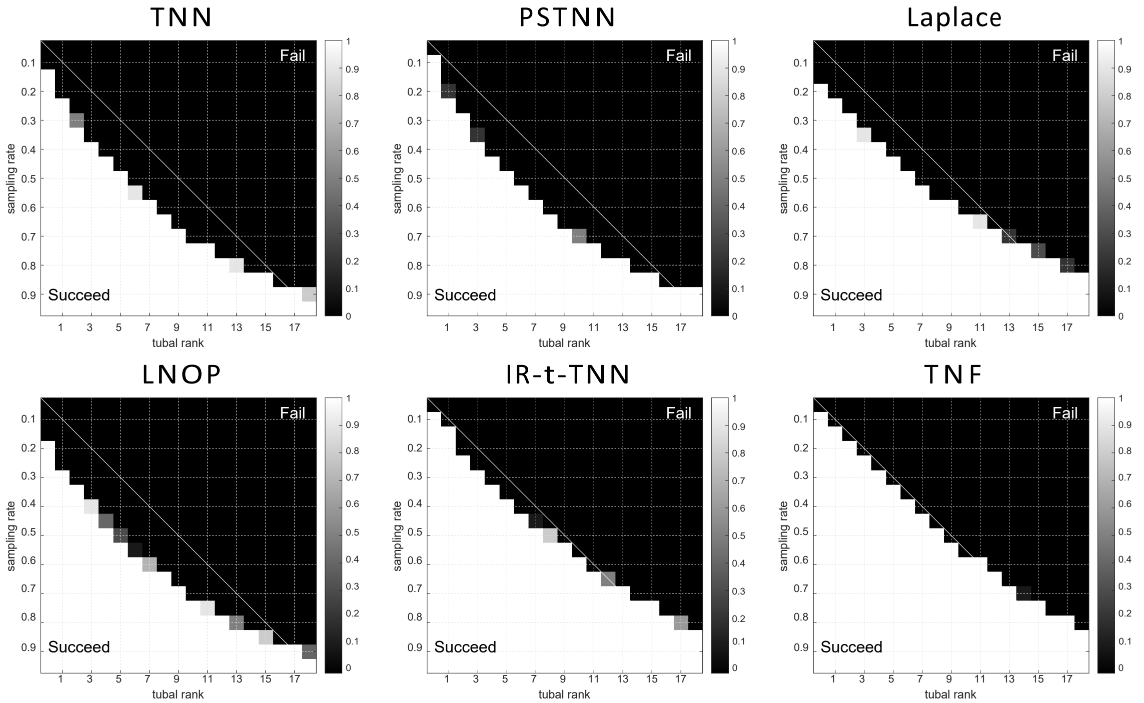

Figure 3. The success rates for the tensor completion problem with varying tubal ranks and sampling rates. Ten independent experiments are conducted, and the percentage of successful recoveries is indicated in each cell. To facilitate comparison, white dashed lines have been added along the diagonal line.

We follow the experimental setup in [17] to test third-order tensors of size 40 s 40 s 20. The tubal rank is from 1 to 19 with an increment of 2, and the sampling rate is from 0 to 0.9 with an increment of 0.1. Based on the parameter sensitivity experiment, we initialize u_1 and u_2 with the value of 10^{-4}, and gradually multiply them by a factor of 1.1 at each outer iteration. Both u_1 and u_2 are capped at a maximum value of 10^{10}.

Figure 3 presents the success rates of tensor completion by various models. Generally, a lower tubal rank or a higher sampling rate leads to a higher success rate. Our method achieves the best recovery results over other competing models (TNN, PSTNN, Laplace, LNOP, and IR-t-TNN), as demonstrated by the smallest black area in Figure 3. Notably, even when the tubal rank is as high as 15, our approach can recover the solution under lower sampling rates much better compared to others.

5.2. Real-world data--video and image inpainting. We conduct experiments on real-world data including videos and color images. We use the peak signal-to-noise ratio (PSNR) [30] and the structural similarity index (SSIM) [47] to quantitatively evaluate the recovery performance.

We start with an inpainting application for videos and images, where we randomly sample only 20\% entries (i.e., 80\% missing values). We consider one video data named ``basketball'''',[1]

[1] https://www1.cs.columbia.edu/CAVE/databases/multispectral

Table 1
Quantitative comparisons of inpainting results obtained by different methods.

Data	Size	Index	Observed	TNN	PSTNN	Laplace	LNOP	IR-t-TNN	TNF
``basketball""	144 s 256 s 40	PSNR	6.2375	21.3773	21.3085	20.5415	21.3804	20.7249	oe.r r ee
		SSIM	0.0722	0.6124	0.6019	0.5630	0.6129	0.5408	o.x ene
``sailboat1""	512 s 768 s 3	PSNR	6.1434	25.0217	24.6523	25.1105	25.0292	24.9662	oe.e ont
		SSIM	0.0692	0.8056	0.7917	0.8075	0.8049	0.8073	o.t oen
``castle""	768 s 512 s 3	PSNR	7.3570	25.7897	25.8843	25.9954	25.8076	ox.e xoe	25.9760
		SSIM	0.0777	0.8412	0.8378	o.t r ee	0.8415	0.8398	0.8438
``window""	512 s 768 s 3	PSNR	13.1233	24.6276	24.7283	22.7186	24.6741	23.7092	or .t r ee
		SSIM	0.6333	0.8383	0.8408	0.8124	0.8390	0.8240	o.t r ee
``sailboat2""	768 s 512 s 3	PSNR	12.5847	28.5425	28.1361	28.2890	28.5608	27.9419	ot .x r ee
		SSIM	0.6514	0.9195	0.9211	0.9160	0.9198	0.9069	o.e oer

for which we set the initial values as $u_1 = 10^{-3}$ and $u_2 = 10^{-6}$. In two color images (``sailboat1"" and ``castle""),[2] we set the initial values as $u_1 = 10^{-3}$ and $u_2 = 10^{-5}$. We fill in the missing values by the proposed TNF model with a comparison to TNN [53], PSTNN [17], Laplace [50], LNOP [5], and IR-t-TNN [46]. The quantitative measures of PSNR and SSIM are reported in Table 1, showcasing that our method achieves the best results. Image recovery results in Figure 4 shows that all the methods yield similar performance visually.

In addition, we investigate two structural missing patterns of letters and grids, each of which is more dificult than randomly missing. In a color image (``window""), we use letters to block off intensities and set initial values as $u_1 = 10^{-6}$ and $u_2 = 10^{-4}$. While in image (``sailboat2""), we use a grid mask and set initial values as $u_1 = 10^{-8}$ and $u_2 = 10^{-5}$. Table 1 and Figure 4 show that the proposed TNF method outperforms the other competing methods.

6. Conclusions.

In this paper, we introduced a novel nonconvex approximation to the tensor tubal rank, referred to as the tensor nuclear over the TNF, and studied low-rank tensor completion problem using the proposed TNF regularization. Among a series of properties of TNF, we studied the local minimum of TNF under a NSP-type condition for a general tensor recovery problem. In addition, we proposed a TNF-based tensor completion model that can be solved eficiently by ADMM with convergence guarantees. We conducted extensive experiments to demonstrate the effectiveness of our model over the state-of-the-art. To ensure computational eficiency, our numerical approach necessitates an inexact scheme for the inner loop, while the current convergence analysis requires the optimal solution for the subproblem. One future direction of this work would be refined to accommodate situations where inexact solutions are employed. Our forthcoming endeavors are geared towards aligning with the methodologies proposed in [6, 26] to establish an enhanced theoretical foundation in this regard. Furthermore, we will analyze the sample size of exact recovery and error bound, considering the challenges posed by the nonconvexity and nonseparability of our proposed model. In addition, an extension of this work to encompass various noise distributions would be an interesting future research direction.

[2] http://r0k.us/graphics/kodak

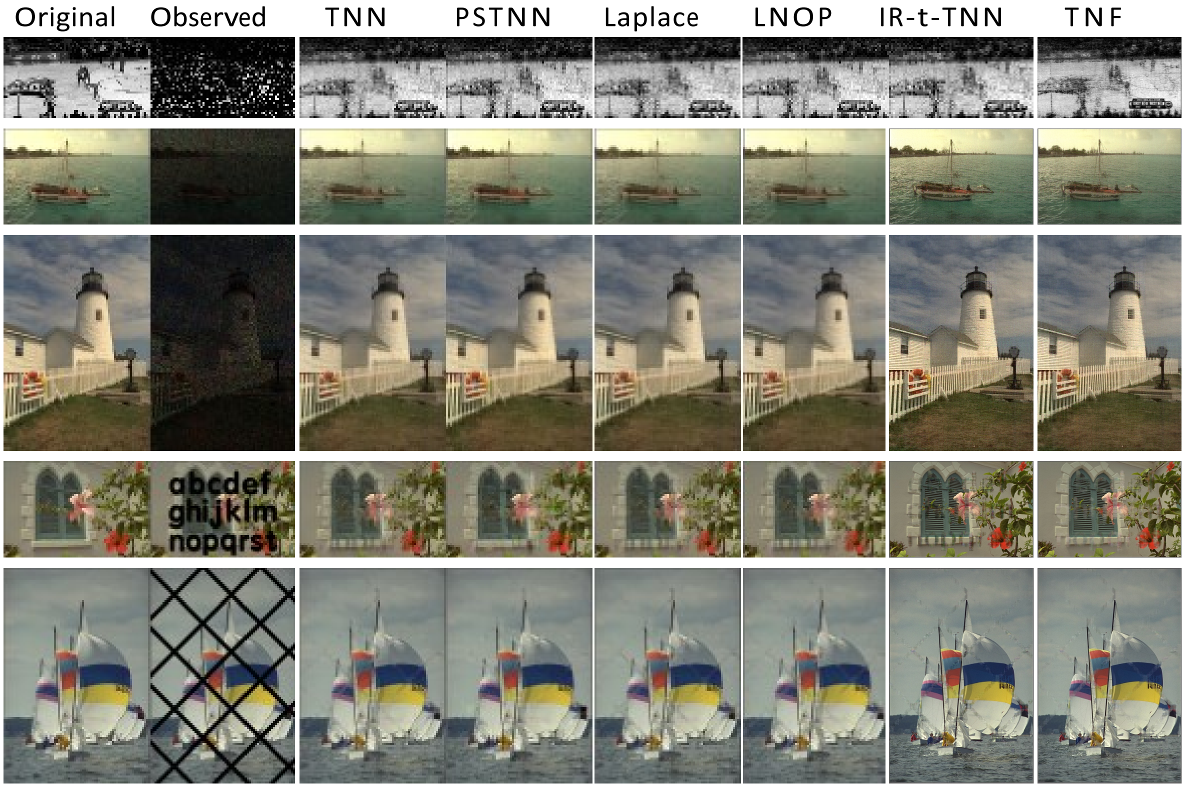

Figure 4. Comparison of video and image inpainting performance on five examples. From left to right: original image, observed input, recovered images by TNN, PSTNN, Laplace, LNOP, IR-t-TNN, and our TNF. From top to bottom: one frame of the video data (``basketball''), color images with random missing data (``sailboat1'' and ``castle''), and color images with structural missing data (``window'' and ``sailboat2'').

Appendix A. Proofs of Theorem 3.5. Given a nonzero low-rank tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with its tubal rank $r \geq n$ that satisfies $\mathcal{F}(\mathcal{X}) = \mathcal{T}$ and a tensor $\mathcal{W} \in \mathcal{N}(\mathcal{F}) \setminus \{\mathcal{O}\}$ with $\|\mathcal{W}\|_F = 1$, we define a function

$$g(t) = \frac{\|\mathcal{X} + t\mathcal{W}\|_t^2}{\|\mathcal{X} + t\mathcal{W}\|_F^2}. \tag{A.1}$$

It is worth noting that g is well-defined as its denominator cannot be zero for any $t \in \mathbb{R}$. Otherwise, it leads to $\mathcal{X} + t\mathcal{W} = \mathcal{O}$, subsequently resulting in $\mathcal{F}(\mathcal{X} + t\mathcal{W}) = \mathcal{O}$. Since $\mathcal{F}(\mathcal{W}) = \mathcal{O}$, we would then arrive at $\mathcal{F}(\mathcal{X}) = \mathcal{O}$, which contradicts the setting that $\mathcal{T} = \mathcal{O}$. In what follows, we examine the numerator of (A.1) and obtain Lemma A.1.

Lemma A.1. For a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r and a tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\|\mathcal{W}\|_F = 1$, it holds for $t \in \mathbb{R}$ that

$$\|\mathcal{X} + t\mathcal{W}\|_t \geq \|\mathcal{X}\|_t + t a_t(\mathcal{W}),$$

where $a_t(\mathcal{W}) := \mathrm{sign}(t)(\|\mathcal{W}_{r^c}\|_t - \|\mathcal{W}_r\|_t)$

Proof. It is straightforward from the TNN definition (2.6) and the triangle inequality that

$$
\begin{aligned}
\|\mathcal{X} + t\mathcal{W}\|_* &= \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{n} a_{ij}(\mathcal{X} + t\mathcal{W}) \\
&\geq \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{n} |a_{ij}(\mathcal{X}) - a_{ij}(t\mathcal{W})| \\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{r} |a_{ij}(\mathcal{X}) - a_{ij}(t\mathcal{W})| + \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=r+1}^{n} |a_{ij}(t\mathcal{W})| \\
&\geq \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{r} (a_{ij}(\mathcal{X}) - a_{ij}(t\mathcal{W})) + \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=r+1}^{n} (a_{ij}(t\mathcal{W})) \\
&= \|\mathcal{X}\|_* - |t|\,\|\mathcal{W}_r\|_* + |t|\,\|\mathcal{W}_{r^c}\|_* \\
&= \|\mathcal{X}\|_* + t\cdot\mathrm{sign}(t)(\|\mathcal{W}_{r^c}\|_* - \|\mathcal{W}_r\|_*).
\end{aligned} \tag{A.2}
$$

∎

It follows from Lemma A.1 that

$$
f(t) := \frac{(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))^2}{\|\mathcal{X} + t\mathcal{W}\|_F^2} \geq g(t) \quad \forall t \in \mathbb{R}.
$$

Note that the function f has a nonzero denominator (same reason as g) and thus is continuous in its domain and differentiable on the intervals $(-\infty, 0)$ and $(0, \infty)$, individually. By some calculations, we deduce the derivative of f for $t > 0$,

$$
\begin{aligned}
f'(t) &= \frac{d}{dt}\left(\frac{(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))^2}{\|\mathcal{X}\|_F^2 + 2t\langle\mathcal{W},\mathcal{X}\rangle + t^2\|\mathcal{W}\|_F^2}\right) \\
&= \frac{2a_*(\mathcal{W})(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))\left(\|\mathcal{X}\|_F^2 + 2t\langle\mathcal{W},\mathcal{X}\rangle + t^2\|\mathcal{W}\|_F^2\right) - \left(2\langle\mathcal{W},\mathcal{X}\rangle + 2t\|\mathcal{W}\|_F^2\right)(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))^2}{(\|\mathcal{X}\|_F^2 + 2t\langle\mathcal{W},\mathcal{X}\rangle + t^2\|\mathcal{W}\|_F^2)^2} \\
&= \frac{2(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))\left[a_*(\mathcal{W})\left(\|\mathcal{X}\|_F^2 + 2t\langle\mathcal{W},\mathcal{X}\rangle + t^2\|\mathcal{W}\|_F^2\right) - \left(\langle\mathcal{W},\mathcal{X}\rangle + t\|\mathcal{W}\|_F^2\right)(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))\right]}{(\|\mathcal{X}\|_F^2 + 2t\langle\mathcal{W},\mathcal{X}\rangle + t^2\|\mathcal{W}\|_F^2)^2} \\
&= \frac{2(\|\mathcal{X}\|_* + t a_*(\mathcal{W}))\left[\left(a_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle\mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_*\right) + \left(a_*(\mathcal{W})\langle\mathcal{W},\mathcal{X}\rangle - \|\mathcal{X}\|_*\|\mathcal{W}\|_F^2\right)t\right]}{(\|\mathcal{X}\|_F^2 + 2t\langle\mathcal{W},\mathcal{X}\rangle + t^2\|\mathcal{W}\|_F^2)^2}.
\end{aligned} \tag{A.3}
$$

The derivative of $f(t)$ for $t < 0$ follows similarly.

Next, we extend the equivalence of the vector norms to tensors in Lemma A.2. The proof is trivial, which is omitted.

Lemma A.2. If $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a tensor of tubal rank r, then

$$
\frac{1}{n_3}\|\mathcal{X}\|_F^2 \leq \|\mathcal{X}\|_*^2 \leq r\|\mathcal{X}\|_F^2.
$$

Lemma A.3. Suppose a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ with tubal rank r that satisfies $\mathcal{F}(\mathcal{X}) = \mathcal{T}$. We define

$$
t_* := \inf_{\mathcal{W},t}\left\{\left|\frac{a_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle\mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_*}{a_*(\mathcal{W})\langle\mathcal{W},\mathcal{X}\rangle - \|\mathcal{X}\|_*\|\mathcal{W}\|_F^2}\right| \;\middle|\; \mathcal{W} \in \mathcal{N}(\mathcal{F})\times\{\mathcal{O}\}, \|\mathcal{W}\|_F = 1, t \neq 0\right\}, \tag{A.4}
$$

where $\alpha_*(\mathcal{W})$ is defined in Lemma A.1. If the linear operator $\mathfrak{F}$ satisfies NSP of order r with a positive constant $s < \frac{\sqrt{r}}{\sqrt{r}+n\sqrt{n_3}}$, then $t_* > 0$.

Proof. For any tensor $\mathcal{W} \in \mathcal{N}(\mathfrak{F}) \setminus \{\mathcal{O}\}$ and $\|\mathcal{W}\|_F = 1$, under the NSP condition (3.8), we get $\|\mathcal{W}_r\|_* \leq s\|\mathcal{W}_{r^c}\|_*$ with $s < \frac{\sqrt{r}}{\sqrt{r}+n\sqrt{n_3}}$ by assumption. Therefore,

$$|\alpha_*(\mathcal{W})| = \|\mathcal{W}_{r^c}\|_* - \|\mathcal{W}_r\|_* \geq \frac{1-s}{s}\|\mathcal{W}_r\|_* \geq \frac{(1-s)r}{sn}\|\mathcal{W}\|_* \geq \frac{(1-s)r}{sn\sqrt{n_3}}, \tag{A.5}$$

where the third inequality is from $\|\mathcal{W}\|_* \geq \frac{1}{\sqrt{n_3}}\|\mathcal{W}\|_F$. For the numerator of (A.4), we get

$$\begin{aligned}
\left| \alpha_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle \mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_* \right| &\geq |\alpha_*(\mathcal{W})|\,\|\mathcal{X}\|_F^2 - |\langle \mathcal{W},\mathcal{X}\rangle|\,\|\mathcal{X}\|_* \\
&\geq |\alpha_*(\mathcal{W})|\,\|\mathcal{X}\|_F^2 - \|\mathcal{W}\|_F\|\mathcal{X}\|_F\|\mathcal{X}\|_* \\
&\geq \frac{(1-s)r}{sn\sqrt{n_3}}\|\mathcal{X}\|_F^2 - \sqrt{r}\,\|\mathcal{X}\|_F^2 \\
&= \frac{\sqrt{r}}{sn\sqrt{n_3}}\|\mathcal{X}\|_F^2\left(\sqrt{r} - s(\sqrt{r} + n\sqrt{n_3})\right) > 0.
\end{aligned} \tag{A.6}$$

As for the denominator of (A.4), we have

$$\begin{aligned}
\left|\alpha_*(\mathcal{W})\langle \mathcal{W},\mathcal{X}\rangle - \|\mathcal{X}\|_*\|\mathcal{W}\|_F^2\right| &\leq |\alpha_*(\mathcal{W})|\,\|\mathcal{W}\|_F\|\mathcal{X}\|_F + \|\mathcal{W}\|_F^2\|\mathcal{X}\|_* \\
&\leq \left|\|\mathcal{W}_r\|_* + \|\mathcal{W}_{r^c}\|_*\right|\|\mathcal{X}\|_F + \|\mathcal{X}\|_* \\
&= \|\mathcal{W}\|_*\|\mathcal{X}\|_F + \|\mathcal{X}\|_* \\
&\leq \sqrt{n}\|\mathcal{X}\|_F + \|\mathcal{X}\|_*.
\end{aligned} \tag{A.7}$$

Combining (A.6) and (A.7), we obtain a nonzero lower bound, i.e.,

$$\frac{\left|\alpha_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle \mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_*\right|}{\left|\alpha_*(\mathcal{W})\langle \mathcal{W},\mathcal{X}\rangle - \|\mathcal{X}\|_*\|\mathcal{W}\|_F^2\right|} \geq \frac{sn\sqrt{n_3}(\sqrt{n}\|\mathcal{X}\|_F + \|\mathcal{X}\|_*)}{\sqrt{r}(\sqrt{r} - s(\sqrt{r} + n\sqrt{n_3}))\|\mathcal{X}\|_F^2} > 0.$$

Hence Lemma A.3 holds. ∎

Now, we are ready to prove Theorem 3.5.

Proof. It follows from the NSP condition (3.8) that

$$t\alpha_*(\mathcal{W}) = t\,\mathrm{sign}(t)(\|\mathcal{W}_{r^c}\|_* - \|\mathcal{W}_r\|_*) = |t|(\|\mathcal{W}_{r^c}\|_* - \|\mathcal{W}_r\|_*) \geq |t|\left(\frac{1}{s} - 1\right)\|\mathcal{W}_r\|_* > 0,$$

which implies $\|\mathcal{X}\|_* + t\alpha_*(\mathcal{W}) > 0\ \forall t \in \mathbb{R}$. By Lemma A.3, there exists a positive number t_* defined in (A.4) such that

$$\left|\alpha_*(\mathcal{W})\langle \mathcal{W},\mathcal{X}\rangle - \|\mathcal{X}\|_*\|\mathcal{W}\|_F^2\right||t| \leq \left|\alpha_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle \mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_*\right|$$

for all $|t| \leq t_*$ and $\mathcal{W} \in \mathcal{N}(\mathfrak{F}) \setminus \{\mathcal{O}\}$ with $\|\mathcal{W}\|_F = 1$. Therefore, we have

$$\begin{aligned}
&\mathrm{sign}[\alpha_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle \mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_* + (\alpha_*(\mathcal{W})\langle \mathcal{W},\mathcal{X}\rangle - \|\mathcal{X}\|_*\|\mathcal{W}\|_F^2)t] \\
&\quad = \mathrm{sign}[\alpha_*(\mathcal{W})\|\mathcal{X}\|_F^2 - \langle \mathcal{W},\mathcal{X}\rangle\|\mathcal{X}\|_*].
\end{aligned} \tag{A.8}$$

From (A.6), $|\partial_t(W)\|X\|_F^2| - |\langle W, X\rangle \|X\|_t| \le 0$, thus leading to

$$\text{sign}[\partial_t(W)\|X\|_F^2 - \langle W, X\rangle \|X\|_t] = \text{sign}(\partial_t(W)) = \text{sign}(t). \tag{A.9}$$

As a result, we have $f^e(t) \le 0$ if $0 < t < t_t$. Then we further compute the subderivative of $f(t)$ at $t = 0$:

$$f_e(0^+) = \lim_{t \to 0^+} \frac{f(t) - f(0)}{t - 0} = \frac{2\|X\|_t\left(\partial_t(W)\|X\|_F^2 - \langle W, X\rangle \|X\|_t\right)}{\|X\|_F^4} \le 0, \tag{A.10}$$

where the inequality is from (A.8) and (A.9). Similarly, we can get $f^e(t) \le 0$ if $-t_t < t < 0$ and $f^e(0^-) \le 0$. Consequently, we have $f(t) \le f(0)$ for any $|t| \le t_t$. Together with $g(t) \le f(t)$, we obtain that $g(t) \le f(0) = g(0)$, which implies

$$\frac{\|X + tW\|_t^2}{\|X + tW\|_F^2} \le \frac{\|X\|_t^2}{\|X\|_F^2} \quad \forall |t| \le t_t.$$

Noticing that t_t is a constant that does not depend on the choice of W, we complete the proof of Theorem 3.5. ■

Appendix B. Proofs of convergence.

To make this paper self-contained, we include some prerequisites from [44] in Lemmas B.1 to B.3.

Lemma B.1 (see [44]). A function $f(x)$ is called strongly convex with parameter u if and only if one of the following conditions holds:

(a) $g(x) = f(x) - \frac{u}{2}\|x\|_2^2$ is convex;
(b) $\langle \partial f(x) - \partial f(y), x - y\rangle \le u\|x - y\|_2^2 \quad \forall x, y$;
(c) $f(y) \le f(x) + \langle \partial f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2 \quad \forall x, y$.

Lemma B.2 (see [44]). The gradient of $f(x)$ is Lipschitz continuous with parameter $L > 0$ if and only if one of the following conditions holds:

(a) $\|\partial f(x) - \partial f(y)\|_2 \le L\|x - y\|_2 \quad \forall x, y$;
(b) $g(x) = \frac{L}{2}\|x\|_2^2 - f(x)$ is convex;
(c) $f(y) \le f(x) + \langle \partial f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2 \quad \forall x, y$.

Lemma B.3 (see [44]). Given a function $f(x) = \frac{1}{\|x\|_2}$ and a set $M_\eta := \{x \,|\, \|x\|_2 \le \eta\}$ for a positive constant $\eta > 0$, we have

$$\|\partial f(x) - \partial f(y)\|_2 \le \frac{2}{\eta^3}\|x - y\|_2 \quad \forall x, y \in M_\eta.$$

Lemma B.4 follows from Lemma B.3 immediately.

Lemma B.4. Given a function $g(X) = \frac{1}{\|X\|_F}$ and a set $M_a := \{X \,|\, \|X\|_F \le a\}$ for a positive constant $a > 0$, we have

$$\|\partial g(X) - \partial g(Y)\|_F \le \frac{2}{a^3}\|X - Y\|_F \quad \forall X, Y \in M_a. \tag{B.1}$$

Proof. Reshaping any tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{M}_a$ into the vector forms of x and y, respectively, and applying Lemma B.3, we have

$$\|\nabla g(\mathcal{X}) - \nabla g(\mathcal{Y})\|_F = \left\| \frac{\mathcal{X}}{\|\mathcal{X}\|_F^3} - \frac{\mathcal{Y}}{\|\mathcal{Y}\|_F^3} \right\|_F = \left\| \frac{x}{\|x\|_2^3} - \frac{y}{\|y\|_2^3} \right\|_2 \le \frac{2}{a^3}\|x - y\|_2 = \frac{2}{a^3}\|\mathcal{X} - \mathcal{Y}\|_F .$$

■

B.1. Proof of Lemma 4.2.

Proof. The optimality condition of the $\mathcal{H}$ subproblem in (4.15) indicates that

$$-\frac{\|\mathcal{X}^{(k+1)}\|_*}{\|\mathcal{H}^{(k+1)}\|_F^3}\mathcal{H}^{(k+1)} - \rho_1\left(\mathcal{X}^{(k+1)} - \mathcal{H}^{(k+1)} + \frac{\mathcal{A}^{(k)}}{\rho_1}\right) = \mathcal{O}, \tag{B.2}$$

where $\mathcal{O} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the zero tensor. Using the dual update $\mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} + \rho_1(\mathcal{X}^{(k+1)} - \mathcal{H}^{(k+1)})$, we have

$$\mathcal{A}^{(k+1)} = -\frac{\|\mathcal{X}^{(k+1)}\|_*}{\|\mathcal{H}^{(k+1)}\|_F^3}\mathcal{H}^{(k+1)}, \tag{B.3}$$

which directly deduces

$$\mathcal{A}^{(k)} = -\frac{\|\mathcal{X}^{(k)}\|_*}{\|\mathcal{H}^{(k)}\|_F^3}\mathcal{H}^{(k)}. \tag{B.4}$$

Then it is straightforward to have

(B.5)
$$\begin{aligned}
\left\|\mathcal{A}^{(k+1)} - \mathcal{A}^{(k)}\right\|_F &= \left\| \frac{\|\mathcal{X}^{(k+1)}\|_*}{\|\mathcal{H}^{(k+1)}\|_F^3}\mathcal{H}^{(k+1)} - \frac{\|\mathcal{X}^{(k)}\|_*}{\|\mathcal{H}^{(k)}\|_F^3}\mathcal{H}^{(k)} \right\|_F \\
&\le \left\| \frac{\mathcal{H}^{(k+1)}}{\|\mathcal{H}^{(k+1)}\|_F^3}\left(\|\mathcal{X}^{(k+1)}\|_* - \|\mathcal{X}^{(k)}\|_* \right) \right\|_F \\
&\quad + \|\mathcal{X}^{(k)}\|_* \left\| \frac{\mathcal{H}^{(k+1)}}{\|\mathcal{H}^{(k+1)}\|_F^3} - \frac{\mathcal{H}^{(k)}}{\|\mathcal{H}^{(k)}\|_F^3} \right\|_F \\
&= \frac{1}{\|\mathcal{H}^{(k+1)}\|_F^2}\left| \|\mathcal{X}^{(k+1)}\|_* - \|\mathcal{X}^{(k)}\|_* \right| + \|\mathcal{X}^{(k)}\|_* \left\| \frac{\mathcal{H}^{(k+1)}}{\|\mathcal{H}^{(k+1)}\|_F^3} - \frac{\mathcal{H}^{(k)}}{\|\mathcal{H}^{(k)}\|_F^3} \right\|_F .
\end{aligned}$$

It follows from Lemma A.2 that $\|\mathcal{A}\|_* \le \sqrt{r}\|\mathcal{A}\|_F \le \sqrt{n}\|\mathcal{A}\|_F \ \forall \mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Therefore,

$$\left| \|\mathcal{X}^{(k+1)}\|_* - \|\mathcal{X}^{(k)}\|_* \right| \le \left\| \mathcal{X}^{(k+1)} - \mathcal{X}^{(k)} \right\|_* \le \sqrt{n}\left\| \mathcal{X}^{(k+1)} - \mathcal{X}^{(k)} \right\|_F . \tag{B.6}$$

According to Lemma B.4 and C1--C2, we have

$$\|\mathcal{X}^{(k)}\|_* \left\| \frac{\mathcal{H}^{(k+1)}}{\|\mathcal{H}^{(k+1)}\|_F^3} - \frac{\mathcal{H}^{(k)}}{\|\mathcal{H}^{(k)}\|_F^3} \right\|_F \le \frac{2M}{\epsilon^3}\left\| \mathcal{H}^{(k+1)} - \mathcal{H}^{(k)} \right\|_F . \tag{B.7}$$

Putting (B.5), (B.6), (B.7) together with Cauchy--Schwarz inequality leads to the desired inequality (4.18). ■

B.2. Proof of Lemma 4.3.

Proof. It is straightforward that the function $L_{T\}C\}}(X, H^{(k)}, A^{(k)}) - I_i(X)$ with fixed $H^{(k)}$ and $A^{(k)}$ is strongly convex with constant u_1. Then Lemma B.1 implies that

$$L_{T\}C\}}(X^{(k+1)}, H^{(k)}, A^{(k)}) - I_i(X^{(k+1)}) \; q \; L_{T\}C\}}(X^{(k)}, H^{(k)}, A^{(k)}) - I_i(X^{(k)}) - \frac{u_1}{2}\left\|X^{(k+1)} - X^{(k)}\right\|_F^2 . \tag{B.8}$$

As $X^{(k+1)}$ and $X^{(k)}$ are the optimal solutions in the X-subproblem in the kth and $(k-1)$th iteration, respectively, they satisfy the constraint, i.e., $I_i(X^{(k+1)}) = I_i(X^{(k)})$, and hence we get

$$L_{T\}C\}}(X^{(k+1)}, H^{(k)}, A^{(k)}) \; q \; L_{T\}C\}}(X^{(k)}, H^{(k)}, A^{(k)}) - \frac{\iota_1}{2}\left\|X^{(k+1)} - X^{(k)}\right\|_F^2 . \tag{B.9}$$

On the other hand, we get

$$\begin{aligned} & L_{T\}C\}}(X^{(k+1)}, H^{(k+1)}, A^{(k)}) - L_{T\}C\}}(X^{(k+1)}, H^{(k)}, A^{(k)}) \\ & = \frac{\|X^{(k+1)}\|_t}{\|H^{(k+1)}\|_F} - \frac{\|X^{(k+1)}\|_t}{\|H^{(k)}\|_F} + \frac{u_1}{2}\left\|X^{(k+1)} - H^{(k+1)}\right\|_F^2 - \frac{u_1}{2}\left\|X^{(k+1)} - H^{(k)}\right\|_F^2 \\ & + \left\langle A^{(k)}, X^{(k+1)} - H^{(k+1)}\right\rangle - \left\langle A^{(k)}, X^{(k+1)} - H^{(k)}\right\rangle . \end{aligned} \tag{B.10}$$

It follows from Lemma B.4 and the assumption C2 that $\frac{1}{\|H\|_F}$ is Lipschitz continuous with parameter $\frac{2}{a^3}$. Using Lemma B.2 (c), we obtain

$$\frac{\|X^{(k+1)}\|_t}{\|H^{(k+1)}\|_F} \; q \; \frac{\|X^{(k+1)}\|_t}{\|H^{(k)}\|_F} - \left\langle \frac{\|X^{(k+1)}\|_t H^{(k)}}{\|H^{(k)}\|_F^3}, H^{(k+1)} - H^{(k)}\right\rangle + \frac{\|X^{(k+1)}\|_t}{a^3}\left\|H^{(k+1)} - H^{(k)}\right\|_F^2 . \tag{B.11}$$

Simple calculations of the third and the fourth terms in (B.10) yield

$$\begin{aligned} & \frac{\iota_1}{2}\left\|X^{(k+1)} - H^{(k+1)}\right\|_F^2 - \frac{\iota_1}{2}\left\|X^{(k+1)} - H^{(k)}\right\|_F^2 \\ & = \frac{u_1}{2}\left\|H^{(k+1)}\right\|_F^2 - \frac{u_1}{2}\left\|H^{(k)}\right\|_F^2 - u_1\left\langle X^{(k+1)}, H^{(k+1)} - H^{(k)}\right\rangle \\ & = \frac{\iota_1}{2}\left\|H^{(k+1)}\right\|_F^2 - \frac{\iota_1}{2}\left\|H^{(k)}\right\|_F^2 - \left\langle u_1 H^{(k+1)} + A^{(k+1)} - A^{(k)}, H^{(k+1)} - H^{(k)}\right\rangle \\ & = -\frac{\iota_1}{2}\left\|H^{(k+1)} - H^{(k)}\right\|_F^2 - \left\langle A^{(k+1)} - A^{(k)}, H^{(k+1)} - H^{(k)}\right\rangle , \end{aligned} \tag{B.12}$$

where the second equality is from the A-update. Putting together (B.10), (B.11), (B.12), we have

(B.13)

$$
\begin{aligned}
&L_{\beta}(X^{(k+1)}, H^{(k+1)}, \Lambda^{(k)}) - L_{\beta}(X^{(k+1)}, H^{(k)}, \Lambda^{(k)}) \\
&\le -\left\langle \frac{\|X^{(k+1)}\|_1}{\|H^{(k)}\|_F^3} H^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle + \frac{\|X^{(k+1)}\|_1}{\epsilon^3} \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 \\
&\quad - \frac{\rho_1}{2} \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 - \left\langle \Lambda^{(k+1)} - \Lambda^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle - \left\langle \Lambda^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle \\
&= \left\langle \frac{\|X^{(k+1)}\|_1}{\|H^{(k+1)}\|_F^3} H^{(k+1)} - \frac{\|X^{(k+1)}\|_1}{\|H^{(k)}\|_F^3} H^{(k)}, H^{(k+1)} - H^{(k)} \right\rangle + \frac{\frac{2}{\epsilon^3}\|X^{(k+1)}\|_1 - \rho_1}{2} \left\| H^{(k+1)} - H^{(k)} \right\|_F^2 \\
&\le -\left(\frac{\rho_1}{2} - \frac{3M}{\epsilon^3} \right) \left\| H^{(k+1)} - H^{(k)} \right\|_F^2,
\end{aligned}
$$

where the equality is from (B.3).

Lastly, the Λ-update in (4.6) leads to

(B.14)

$$
\begin{aligned}
&L_{\beta}(X^{(k+1)}, H^{(k+1)}, \Lambda^{(k+1)}) - L_{\beta}(X^{(k+1)}, H^{(k+1)}, \Lambda^{(k)}) \\
&= \left\langle \Lambda^{(k+1)}, X^{(k+1)} - H^{(k+1)} \right\rangle - \left\langle \Lambda^{(k)}, X^{(k+1)} - H^{(k+1)} \right\rangle \\
&= \left\langle \Lambda^{(k+1)} - \Lambda^{(k)}, \frac{1}{\rho_1}(\Lambda^{(k+1)} - \Lambda^{(k)}) \right\rangle = \frac{1}{\rho_1} \|\Lambda^{(k+1)} - \Lambda^{(k)}\|_F^2 \\
&\le \frac{2n}{\rho_1 \epsilon^4} \left\| X^{(k+1)} - X^{(k)} \right\|_F^2 + \frac{4M^2}{\rho_1 \epsilon^6} \left\| H^{(k+1)} - H^{(k)} \right\|_F^2.
\end{aligned}
$$

By putting together (B.9), (B.13), (B.14) and using the Cauchy--Schwarz inequality, we get

(B.15)

$$
\begin{aligned}
L_{\beta}\left(X^{(k+1)}, H^{(k+1)}, \Lambda^{(k+1)} \right) &\le L_{\beta}\left(X^{(k)}, H^{(k)}, \Lambda^{(k)} \right) \\
&\quad - c_1 \left\| X^{(k+1)} - X^{(k)} \right\|_F^2 - c_2 \left\| H^{(k+1)} - H^{(k)} \right\|_F^2,
\end{aligned}
$$

where $c_1 = \frac{\rho_1}{2} - \frac{2n}{\rho_1 \epsilon^4}$ and $c_2 = \frac{\rho_1}{2} - \frac{3M}{\epsilon^3} - \frac{4M^2}{\rho_1 \epsilon^6}$. If ρ_1 is suficiently large, we can guarantee that c_1, c_2 are strictly positive. ∎

B.3. Proof of Lemma 4.4.

Proof. According to the optimal condition of X in (4.6), there exist $P^{(k+1)} \in \partial(\|X^{(k+1)}\|_1)$ and $T^{(k+1)} \in \partial I_{\Omega}(X^{(k+1)})$ such that

(B.16)
$$
\frac{P^{(k+1)}}{\|H^{(k)}\|_F} + T^{(k+1)} + \rho_1(X^{(k+1)} - H^{(k)}) + \Lambda^{(k)} = 0.
$$

Now, we denote

(B.17)
$$
W_1^{(k+1)} := \frac{P^{(k+1)}}{\|H^{(k+1)}\|_F} + T^{(k+1)} + \rho_1\left(X^{(k+1)} - H^{(k+1)} \right) + \Lambda^{(k+1)}.
$$

By the definition of subgradient, it is straightforward that

$$
W_1^{(k+1)} \in \partial_X L_{\beta}(X^{(k+1)}, H^{(k+1)}, \Lambda^{(k+1)}).
$$

Putting (B.16) and (B.17) together, we have

$$W_1^{(k+1)} = \left(\frac{1}{\|H^{(k+1)}\|_F} - \frac{1}{\|H^{(k)}\|_F}\right) P^{(k+1)} + \mu_1(-H^{(k+1)} + H^{(k)}) + A^{(k+1)} - A^{(k)}. \tag{B.18}$$

Note that

$$\partial(\|A\|_{\circledast}) = \{U * V^* + J \mid U^* * J = 0, J * V = 0, \|J\| \le 1\}, \tag{B.19}$$

where $A = U * S * V^*$.

Additionally, for any tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have

$$\|A\|_F = \frac{1}{\sqrt{n_3}}\|\bar{A}\|_F \le \sqrt{\frac{\operatorname{rank}(\bar{A})}{n_3}}\|\bar{A}\|_2 \le \sqrt{n}\|A\|. \tag{B.20}$$

Based on (B.19) and (B.20), we do the skinny t-SVD of $X^{(k+1)}$ by $X^{(k+1)} = U^{(k+1)} * S^{(k+1)} * (V^{(k+1)})^*$; then we get

$$\begin{aligned} \|P^{(k+1)}\|_F &= \|U^{(k+1)} * (V^{(k+1)})^* + J^{(k+1)}\|_F \\ &\le \sqrt{n}\|U^{(k+1)} * (V^{(k+1)})^* + J^{(k+1)}\| \\ &\le \sqrt{n}\|U^{(k+1)} * (V^{(k+1)})^*\| + \sqrt{n}\|J^{(k+1)}\| \\ &\le 2\sqrt{n} \le 2\sqrt{n_1}. \end{aligned} \tag{B.21}$$

Considering that $\|H^{(k)}\|_F$ and $\|H^{(k+1)}\|_F$ are lower bounded by a, we can estimate an upper bound of $\|W_1^{(k+1)}\|_F$,

$$\begin{aligned} \left\|W_1^{(k+1)}\right\|_F &\le \left|\frac{1}{\|H^{(k+1)}\|_F} - \frac{1}{\|H^{(k)}\|_F}\right| \left\|P^{(k+1)}\right\|_F + \mu_1\left\|H^{(k+1)} - H^{(k)}\right\|_F + \left\|A^{(k+1)} - A^{(k)}\right\|_F \\ &= \left|\frac{\|H^{(k+1)}\|_F - \|H^{(k)}\|_F}{\|H^{(k+1)}\|_F\|H^{(k)}\|_F}\right| \left\|P^{(k+1)}\right\|_F + \mu_1\left\|H^{(k+1)} - H^{(k)}\right\|_F + \left\|A^{(k+1)} - A^{(k)}\right\|_F \\ &\le \frac{2\sqrt{n_1}}{a^2}\left\|H^{(k+1)} - H^{(k)}\right\|_F + \mu_1\left\|H^{(k+1)} - H^{(k)}\right\|_F + \left\|A^{(k+1)} - A^{(k)}\right\|_F \\ &= \left(\frac{2\sqrt{n_1}}{a^2} + \mu_1\right)\left\|H^{(k+1)} - H^{(k)}\right\|_F + \left\|A^{(k+1)} - A^{(k)}\right\|_F. \end{aligned} \tag{B.22}$$

Denote two tensors as follows:

$$\begin{aligned} W_2^{(k+1)} &:= -\frac{\|X^{(k+1)}\|_{\circledast}}{\|H^{(k+1)}\|_F^3} H^{(k+1)} - \mu_1\left(X^{(k+1)} - H^{(k+1)}\right) - A^{(k+1)}, \\ W_3^{(k+1)} &:= X^{(k+1)} - H^{(k+1)}. \end{aligned} \tag{B.23}$$

By (B.2) and the A-update in (4.6), we get

$$W_2^{(k+1)} = A^{(k)} - A^{(k+1)} \quad \text{and} \quad W_3^{(k+1)} = \frac{1}{\iota_1}\left(A^{(k+1)} - A^{(k)}\right). \tag{B.24}$$

Let $W^{(k+1)} = (W_1^{(k+1)}, W_2^{(k+1)}, W_3^{(k+1)})$; we get $W^{(k+1)} \in \partial L_{T\}C\}}\left(X^{(k+1)}, H^{(k+1)}, A^{(k+1)}\right)$ and $\|W^{(k+1)}\|_F^2 = \|W_1^{(k+1)}\|_F^2 + \|W_2^{(k+1)}\|_F^2 + \|W_3^{(k+1)}\|_F^2$

$$\begin{aligned} &\leq 2\left(\frac{2\sqrt{n_1}}{a^2} + u_1\right)^2 \left\|H^{(k+1)} - H^{(k)}\right\|_F^2 + \left(3 + \frac{1}{u_1^2}\right)\left\|A^{(k+1)} - A^{(k)}\right\|_F^2 \\ &\leq a_1 \left\|X^{(k+1)} - X^{(k)}\right\|_F^2 + a_2 \left\|H^{(k+1)} - H^{(k)}\right\|_F^2, \end{aligned} \tag{B.25}$$

where $a_1 = \frac{2n_1 n_2 (3u_1^2+1)}{u_1^2 a^4}$, $a_2 = \frac{16 n_1}{\epsilon^4} + 4u_1^2 + \frac{4M^2(3u_1^2+1)}{u_1^2 a^6}$. ■

B.4. Proof of Theorem 4.5.

Proof.

(i) We first show $\{A^{(k)}\}$ is bounded. From (B.4), we have

$$\|A^{(k)}\|_F = \left\| -\frac{X^{(k)}}{\|H^{(k)}\|_F^3} H^{(k)} \right\|_F = \frac{\|X^{(k)}\|_1}{\|H^{(k)}\|_F^2},$$

which suggests that $\{A^{(k)}\}$ is bounded under assumptions C1--C2. Therefore, $\{H^{(k)}\}$ is also bounded due to the H-update in (4.17).

(ii) It follows from (B.15) that

$$\begin{aligned} L_{T\}C\}}\left(X^{(k)}, H^{(k)}, A^{(k)}\right) \leq\ & L_{T\}C\}}\left(X^{(0)}, H^{(0)}, A^{(0)}\right) \\ & - c_1 \sum_{j=0}^{k-1} \left\|X^{(j+1)} - X^{(j)}\right\|_F^2 - c_2 \sum_{j=0}^{k-1} \left\|H^{(j+1)} - H^{(j)}\right\|_F^2. \end{aligned} \tag{B.26}$$

Using (4.5), we obtain a lower bound of

$$\begin{aligned} & L_{T\}C\}}\left(X^{(k)}, H^{(k)}, A^{(k)}\right) \\ &= \frac{\|X^{(k)}\|_t}{\|H^{(k)}\|_F} + I_i(X^{(k)}) + \frac{u_1}{2}\left\|X^{(k)} - H^{(k)}\right\|_F^2 + \left\langle A^{(k)}, X^{(k)} - H^{(k)} \right\rangle \\ &= \frac{\|X^{(k)}\|_t}{\|H^{(k)}\|_F} + I_i(X^{(k)}) + \frac{u_1}{2}\left\|X^{(k)} - H^{(k)} + \frac{1}{u_1}A^{(k)}\right\|_F^2 - \frac{1}{2\iota_1}\left\|A^{(k)}\right\|_F^2 \\ &\geq -\frac{1}{2u_1}\left\|A^{(k)}\right\|_F^2, \end{aligned}$$

due to the boundness of $\{A^{(k)}\}$. Let $k \to \infty$; by $L_{T\}C\}}(X^{(k)}, H^{(k)}, A^{(k)})$ having a lower bound, we know that $\sum_{j=0}^{k} \|X^{(j+1)} - X^{(j)}\|_F^2$ and $\sum_{j=0}^{k} \|H^{(j+1)} - H^{(j)}\|_F^2$ are finite, which implies that $\|X^{(k+1)} - X^{(k)}\|_F^2 \to 0$ and $\|H^{(k+1)} - H^{(k)}\|_F^2 \to 0$ as $k \to \infty$. Then we can get $\|A^{(k+1)} - A^{(k)}\|_F^2 \to 0$ due to (4.18).

(iii) Since $\{\mathcal{X}^{(k)},\mathcal{H}^{(k)},\mathcal{A}^{(k)}\}$ is bounded, the Bolzano--Weierstrass theorem suggests that there exists a convergent subsequence, defined by $\{\mathcal{X}^{(k_i)},\mathcal{H}^{(k_i)},\mathcal{A}^{(k_i)}\} \to \{\mathcal{X}^t,\mathcal{H}^t,\mathcal{A}^t\}$. It further follows from (ii) that $\{\mathcal{X}^{(k_i+1)},\mathcal{H}^{(k_i+1)},\mathcal{A}^{(k_i+1)}\} \to \{\mathcal{X}^t,\mathcal{H}^t,\mathcal{A}^t\}$, which implies that $\|\mathcal{H}^{(k_i+1)} - \mathcal{H}^{(k_i)}\|_F \to 0$ and $\|\mathcal{A}^{(k_i+1)} - \mathcal{A}^{(k_i)}\|_F \to 0$. Hence Lemma 4.4 guarantees that the zero tensor $\mathcal{O} \in \partial L_{T,C}(\mathcal{X}^t,\mathcal{H}^t,\mathcal{A}^t)$. ■

REFERENCES

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality, Math. Oper. Res., 35 (2010), pp. 438--457.

[2] C. F. Beckmann and S. M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis, Neuroimage, 25 (2005), pp. 294--311.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn., 3 (2011), pp. 1--122.

[4] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, and Z. Xu, Total variation regular-ized tensor RPCA for background subtraction from compressive measurements, IEEE Trans. Image Process., 25 (2016), pp. 4075--4090.

[5] L. Chen, X. Jiang, X. Liu, and Z. Zhou, Robust low-rank tensor recovery via nonconvex singular value minimization, IEEE Trans. Image Process., 29 (2020), pp. 9044--9059.

[6] L. Chen, X. Li, D. Sun, and K.-C. Toh, On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming, Math. Program., 185 (2021), pp. 111--161.

[7] A. Cohen, W. Dahmen, and R. DeVore, Compressed sensing and best k-term approximation, J. Amer. Math. Soc., 22 (2009), pp. 211--231.

[8] D. L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via L_1 minimization, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 2197--2202.

[9] G. Ely, S. Aeron, N. Hao, and M. E. Kilmer, 5D seismic data completion and denoising using a novel class of tensor decompositions, Geophysics, 80 (2015), pp. V83--V95.

[10] H. Foucart and R. Simon, A Mathematical Introduction to Compressive Sensing, Birkhäuser/Springer, New York, 2013.

[11] S. Friedland and L.-H. Lim, Nuclear norm of higher-order tensors, Math. Comp., 87 (2018), pp. 1255--1281.

[12] E. Frolov and I. Oseledets, Tensor methods and recommender systems, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov., 7 (2017), e1201.

[13] S. Gandy, B. Recht, and I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, Inverse Problems, 27 (2011), 025010.

[14] L. He, X. Kong, P. S. Yu, X. Yang, A. B. Ragin, and Z. Hao, DuSK: A dual structure-preserving kernel for supervised tensor learning with applications to neuroimages, in Proceedings of the 2014 SIAM International Conference on Data Mining, SIAM, Philadelphia, 2014, pp. 127--135, https://doi.org/10.1137/1.9781611973440.15.

[15] J. C. Ho, J. Ghosh, and J. Sun, Marble: High-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization, in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 115--124.

[16] V. N. Ioannidis, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, Coupled graphs and tensor factorization for recommender systems and community detection, IEEE Trans. Knowl. Data Eng., 33 (2019), pp. 909--920.

[17] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, Multi-dimensional imaging data recov-ery via minimizing the partial sum of tubal nuclear norm, J. Comput. Appl. Math., 372 (2020), 112680.

[18] W. Jiang, J. Zhang, C. Zhang, L. Wang, and H. Qi, Robust low tubal rank tensor completion via factor tensor norm minimization, Pattern Recognit., 135 (2023), 109169.

[19] H. A. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemom., 14 (2000), pp. 105--122.
[20] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 148--172, https://doi.org/10.1137/110837711.
[21] M. E. Kilmer and C. D. Martin, Factorization strategies for third-order tensors, Linear Algebra Appl., 435 (2011), pp. 641--658.
[22] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM Rev., 51 (2009), pp. 455--500, https://doi.org/10.1137/07070111X.
[23] H. Kong, X. Xie, and Z. Lin, t-Schatten-p norm for low-rank tensor recovery, IEEE J. Sel. Top. Signal. Process., 12 (2018), pp. 1405--1419.
[24] G. Li and T. K. Pong, Global convergence of splitting methods for nonconvex composite optimization, SIAM J. Optim., 25 (2015), pp. 2434--2460, https://doi.org/10.1137/140998135.
[25] X. Li, Y. Ye, and X. Xu, Low-rank tensor completion with total variation for visual data inpainting, in Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, 2017, https://doi.org/10.1609/aaai.v31i1.10776.
[26] L. Liang, D. Sun, and K.-C. Toh, An inexact augmented Lagrangian method for second-order cone programming with applications, SIAM J. Optim., 31 (2021), pp. 1748--1773, https://doi.org/10.1137/20M1374262.
[27] J. Liu, P. Musialski, P. Wonka, and J. Ye, Tensor completion for estimating missing values in visual data, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2012), pp. 208--220.
[28] Y. Liu, Tensors for Data Processing: Theory, Methods, and Applications, Academic Press, 2021.
[29] C. Lu, Transforms based tensor robust PCA: Corrupted low-rank tensors recovery via convex optimization, in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 1125--1132, https://doi.org/10.1109/ICCV48922.2021.00118.
[30] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, IEEE Trans. Pattern Anal. Mach. Intell., 42 (2019), pp. 925--938.
[31] S. Miron, Y. Zniyed, R. Boyer, A. Lima Ferrer de Almeida, G. Favier, D. Brie, and P. Comon, Tensor methods for multisensor signal processing, IET Signal Process., 14 (2020), pp. 693--709.
[32] M. Me}rup, L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred, Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG, NeuroImage, 29 (2006), pp. 938--947.
[33] D. Muti and S. Bourennane, Multidimensional filtering based on a tensor approach, Signal Process., 85 (2005), pp. 2338--2353.
[34] J. Popa, S. E. Minkoff, and Y. Lou, An improved seismic data completion algorithm using low-rank tensor optimization: Cost reduction and optimal data orientation, Geophysics, 86 (2021), pp. V219--V232.
[35] Y. Qiu, G. Zhou, Q. Zhao, and S. Xie, Noisy tensor completion via low-rank tensor ring, IEEE Trans. Neural Netw. Learn. Syst., 35 (2024), pp. 1127--1141.
[36] Y. Rahimi, C. Wang, H. Dong, and Y. Lou, A scale-invariant approach for sparse signal recovery, SIAM J. Sci. Comput., 41 (2019), pp. A3649--A3672, https://doi.org/10.1137/18M123147X.
[37] B. Romera-Paredes and M. Pontil, A new convex relaxation for tensor completion, Adv. Neural Inf. Process. Syst., 26 (2013), pp. 2967--2975.
[38] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, Tensor-based formulation and nuclear norm regularization for multienergy computed tomography, IEEE Trans. Image Process., 23 (2014), pp. 1678--1693.
[39] A. M. Tillmann and M. E. Pfetsch, The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing, IEEE Trans. Inform. Theory, 60 (2013), pp. 1248--1259.
[40] H. Tran and C. Webster, A class of null space conditions for sparse recovery via nonconvex, non-separable minimizations, Results Appl. Math., 3 (2019), 100011.
[41] L. R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika, 31 (1966), pp. 279--311.

[42] M. A. Vasilescu and D. Terzopoulos, Multilinear subspace analysis of image ensembles, in Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, IEEE, 2003, pp. II--93.

[43] C. Wang, M. Tao, C.-N. Chuah, J. Nagy, and Y. Lou, Minimizing L_1 over L_2 norms on the gradient, Inverse Problems, 38 (2022), 065011.

[44] C. Wang, M. Tao, J. G. Nagy, and Y. Lou, Limited-angle CT reconstruction via the L_1/L_2 minimization, SIAM J. Imaging Sci., 14 (2021), pp. 749--777, https://doi.org/10.1137/20M1341490.

[45] C. Wang, M. Yan, Y. Rahimi, and Y. Lou, Accelerated schemes for the L_1/L_2 minimization, IEEE Trans. Signal Process., 68 (2020), pp. 2660--2669.

[46] H. Wang, F. Zhang, J. Wang, T. Huang, J. Huang, and X. Liu, Generalized nonconvex approach for low-tubal-rank tensor recovery, IEEE Trans. Neural Netw. Learn. Syst., 33 (2022), pp. 3305--3319.

[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process., 13 (2004), pp. 600--612.

[48] Y. Wu, D. Chen, C. Li, and N. Tang, Bayesian tensor logistic regression with applications to neuroimaging data analysis of Alzheimer's disease, Stat. Methods Med. Res., 31 (2022), pp. 2368--2382.

[49] S. Xia, D. Qiu, and X. Zhang, Tensor Factorization via Transformed Tensor-Tensor Product for Image Alignment, preprint, https://arxiv.org/abs/2212.05719, 2022.

[50] W. Xu, X. Zhao, T. Ji, J. Miao, T. Ma, S. Wang, and T. Huang, Laplace function based nonconvex surrogate for low-rank tensor completion, Signal Process.: Image Commun., 73 (2019), pp. 62--69.

[51] M. Yang, Q. Luo, W. Li, and M. Xiao, Multiview clustering of images with tensor rank minimization via nonconvex approach, SIAM J. Imaging Sci., 13 (2020), pp. 2361--2392, https://doi.org/10.1137/20M1318006.

[52] J. Ye, L. Wang, G. Li, D. Chen, S. Zhe, X. Chu, and Z. Xu, Learning compact recurrent neural networks with block-term tensor decomposition, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9378--9387.

[53] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3842--3849.

[54] X. Zhao, M. Bai, D. Sun, and L. Zheng, Robust tensor completion: Equivalent surrogates, error bounds, and algorithms, SIAM J. Imaging Sci., 15 (2022), pp. 625--669, https://doi.org/10.1137/21M1429539.

[55] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-Y. Ji, and T.-H. Ma, Tensor N-tubal rank and its convex relaxation for low-rank tensor recovery, Inform. Sci., 532 (2020), pp. 170--189.