t-Schatten-p Norm for Low-Rank Tensor Recovery

Hao Kong, Xingyu Xie, *Student Member, IEEE*, and Zhouchen Lin, *Fellow, IEEE*

***Abstract*—In this paper, we propose a new definition of tensor Schatten-p norm (t-Schatten-p norm) based on t-SVD, and prove that this norm has similar properties to matrix Schatten-p norm. More importantly, the t-Schatten-p norm can better approximate the ℓ_1 norm of the tensor multi-rank with $0 < p < 1$. Therefore, it can be used for the Low-Rank Tensor Recovery problems as a tighter regularizer. We further prove the tensor multi-Schatten-p norm surrogate theorem and give an efficient algorithm accordingly. By decomposing the target tensor into many small-scale tensors, the non-convex optimization problem $(0 < p < 1)$ is transformed into many convex sub-problems equivalently, which can greatly improve the computational efficiency when dealing with large-scale tensors. Finally, we provide the theories on the conditions for exact recovery in the noiseless case and give the corresponding error bounds for the noise case. Experimental results on both synthetic and real-world datasets demonstrate the superiority of our t-Schattern-p norm in the Tensor Robust Principle Component Analysisand the Tensor Completion problems.**

***Index Terms*—Tensor Schatten-p norm, low-rank, tensor decomposition, convex optimization.**

I. Introduction

IN COMPUTER vision and pattern recognition, data structures are becoming more and more complex. Thus multi-dimensional arrays (also called as tensors) attract more and more attention recently. Many problems can be converted to the Low-Rank Tensor Recovery (LRTR) problems, such as video denoising [3], video inpainting [4], subspace clustering [5], recommendation systems [6], multitask learning [7], etc. The LRTR problem aims to recover the original low-rank tensor based on the observed corrupted/disturbed tensor. It can be formulated as the following problem:

$$\min_{\mathcal{X}} \ \operatorname{rank}(\mathcal{X}), \quad \text{s.t. } \Psi(\mathcal{X}) = \mathcal{T}, \tag{1}$$

where $\mathcal{T}$ is the observed measurement by a linear operator $\Psi(\cdot)$ and $\mathcal{X}$ is the clean data. Similar to the matrix case, the operation $\operatorname{rank}(\cdot)$ works as a sparsity regularization of tensor singular values. Unfortunately, none of the existing definitions of tensor rank work well in practice. They are all related to particular tensor decompositions [8]. For example, CP-rank [9] is based on the CANDECOMP/PARAFAC decomposition [10]; Tucker-n-rank [11] is based on the Tucker Decomposition [12]; and tensor multi-rank and tubal-rank [2] are based on t-SVD [1]. Minimizing the rank function directly is usually NP-hard and is difficult to be solved within polynomial time, hence we often replace the function $\operatorname{rank}(\mathcal{X})$ by its convex/non-convex surrogate function $f(\mathcal{X})$:

$$\min_{\mathcal{X}} \ f(\mathcal{X}), \quad \text{s.t. } \Psi(\mathcal{X}) = \mathcal{T}. \tag{2}$$

The main difference among the present LRTR models is the choice of surrogate function $f(\cdot)$. Based on different definitions of tensor singular values, various tensor nuclear norms or tensor Schatten-p norms[1] are proposed as the rank surrogates [1], [13], [14]. But they all have some limitations when applied to real problems.

Based on CP-decomposition [10], Friedland *et al.* [14] introduce cTNN (Tensor Nuclear Norm based on CP) as the convex relaxation of the tensor CP-rank:

$$\|\mathcal{T}\|_{cTNN} = \inf\left\{\sum_{i=1}^{r}|\lambda_i| : \mathcal{T} = \sum_{i=1}^{r}\lambda_i \mathbf{u}_i \circ \mathbf{v}_i \circ \mathbf{w}_i\right\}, \tag{3}$$

where $\|\mathbf{u}_i\| = \|\mathbf{v}_i\| = \|\mathbf{w}_i\| = 1$ and $\circ$ represents the vector outer product. Yuan *et al.* [15] give the sub-gradient of cTNN by leveraging its dual property, therefore we can solve the cTNN minimization problem by using some traditional gradient-based methods. It is worth mentioning that, for a given tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, calculating its CP-rank [9] is usually NP-complete [16], [17], which means that we cannot verify the consistency of cTNN's implicit decomposition with the ground-truth CP-decomposition. Moreover, it is hard to measure the cTNN's tightness relative to the CP-rank since whether cTNN satisfies the continuous analogue of Comon's conjecture [14] remains unknown. What's more, inconsistent with the

[1]Schattern-p norm is only a pseudo-norm when $0 < p < 1$.

Manuscript received April 15, 2018; revised September 17, 2018 and October 20, 2018; accepted October 25, 2018. Date of publication November 1, 2018; date of current version December 17, 2018. The work of Z. Linwas supported in part by National Basic Research Program of China (973 Program) under Grant 2015CB352502,in part by National Natural Science Foundation of China under Grants 61625301 and 61731018, and in part by Microsoft Research Asia. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Rene Vidal. *(Corresponding author: Zhouchen Lin.)*

H. Kong and Z. Lin are with the Key Laboratory of Machine Perception (MoE), School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: konghao@pku.edu.cn; zlin@pku.edu.cn).

X. Xie is with the College of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: nuaaxing@gmail.com).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author. The material includes a Supplemental Material of our proofs. Contact konghao@pku.edu.cn for further questions about this work.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2018.2879185

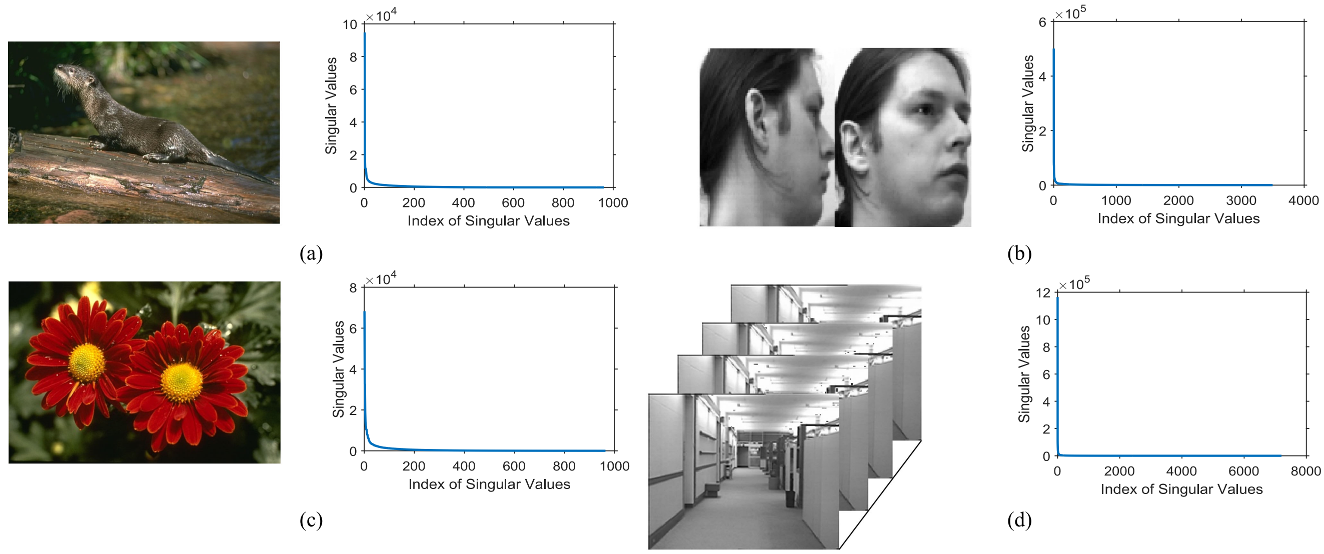


Fig. 1. Illustrations of the low tubal-rank properties of some datasets. (a) and (c) are from the Berkeley Segmentation dataset. (b) is from the UMist Faces dataset. (d) is from the YUV Video Sequences dataset.

two-dimensional case, one cannot extent cTNN to the tensor Schatten-p norm because the infimum will be identically 0 when the ℓ_1 norm of the coefficients is replaced by an ℓ_p norm for any $p > 1$ [14]. All the above reasons limit the application of cTNN to the LRTR problem.

To avoid the NP-complete CP decomposition, Liu *et al.* [13] define a kind of tensor nuclear norm named SNN (Sum of Nuclear Norm) based on the Tucker decomposition [12]:

$$\|\mathcal{T}\|_{SNN} = \sum_{i=1}^{dim} \left\|\mathbf{T}_{(i)}\right\|_{*}, \tag{4}$$

where $\mathbf{T}_{(i)}$ denotes unfolding the tensor along the i-th dimension and $\|\cdot\|_*$ is the nuclear norm of a matrix, i.e., sum of singular values. Because SNN is easy to compute, it has been widely used, e.g., [13], [18], [19]. By considering the subspace structure in each mode, Kasai *et al.* [18] propose a Riemannian manifold based tensor completion method (RMTC). However, Paredes *et al.* [20] point out that SNN is not the tightest convex relaxation of the Tucker-n-rank [11], and is actually an overlap regularization of it. They further propose an alternative convex relaxation of Tucker-n-rank [11] which is tighter than SNN. Tomioka *et al.* [21] introduce a tensor Schatten-p norm based on SNN, and Li *et al.* [19] achieve a better experimental results on the tensor completion (TC) problem than the original SNN. Nonetheless, they still simply unfold the high-order tensor into matrices, which will unavoidably destroy the intrinsic structure of tensor data.

In order to maintain the internal structure of high-dimensional arrays, Kilmer *et al.* [1] propose a new tensor decomposition named t-SVD. Zhang *et al.* [4] give a definition of the nuclear norm corresponding to t-SVD, i.e., Tensor Nuclear Norm (TNN). Further more, they point out that TNN is the tightest convex relaxation to ℓ_1 norm of the tensor multi-rank[2] within the unit ball of the tensor spectral norm.[3]

When arranging image or video data into matrices [13], they often lie on a union of low-rank subspaces. Fortunately, the original tensor data also have a low multi-rank (or tubal-rank) structures. Fig. 1 shows the singular values of all frontal slices of several commonly used datasets. It is easy to see that most singular values are very close to 0 and much smaller than the largest ones. So the related problems can be solved effectively by t-SVD based low-rank methods. By adopting TNN, [3] and [22] propose the exact recovery conditions of TRPCA and TC problems, respectively.

Due to consdering the internal structure of data, TNN has been widely used in recent years. Nevertheless, when dealing with large-scale tensor data, the computational complexity of TNN grows dramatically. For instance, when solving TC problems by TNN, the computational complexity at each iteration is $\mathcal{O}(n_1 n_2 n_3(\log n_3 + \min\{n_1, n_2\}))$, which consumes several hours to complete a tensor with size $500 \times 500 \times 500$. To avoid this high complexity, Zhou *et al.* [23] and Liu *et al.* [24] utilize the tensor factorization method to preserve the low-rank structure, and they only maintain two smaller tensors during each optimization iteration. By decomposing a large-scale tensor into two skinny ones, the computational cost at each iteration drops to $\mathcal{O}(r(n_1 + n_2)n_3 \log n_3 + r n_1 n_2 n_3))$ [23]. Although the complexity is reduced, their methods do not consider the balance between factors, hence they cannot prevent the extremely imbalanced tensor decompositions, which will make their obtained tensors violate the incoherence conditions. Note that incoherence is the essential condition for successful completion.

[2]The tensor multi-rank is a vector with each entry representing the rank of a frontal slice after Fourier transform along the third dimension.

[3]The related definition of the tensor spectral norm can be found in [4].

To break the limits of existing methods, in this paper we propose a new tensor Schatten-p norm (t-Schatten-p norm) which is defined in Eq. (11) based on t-product [1]. The proposed norm is of similar properties to matrix Schatten-p norm. Additionally, when $0 < p < 1$ this Schatten-p norm is a tighter regularizer than TNN to approximate the ℓ_1 norm of tensor multi-rank [2]. Furthermore, inspired by [23] and [25], we extend the matrix norm surrogate theorem to the tensor case. By using the new theorem and t-product, when $0 < p < 1$ we decompose the target tensor $\mathcal{T}$ into many small-scale tensors $\{\mathcal{T}_i\}$ with $\mathcal{T} = \mathcal{T}_1 * \cdots * \mathcal{T}_I$, and then we minimize the weighted sum of convex Schatten-p norms $\sum_i \frac{1}{p_i} \|\mathcal{T}_i\|_{S_{p_i}}^{p_i}$, where $p_i \geq 1, \forall i$. In this way, the original non-convex non-smooth optimization problem is divided into many convex sub-problems. Hence we not only reduce the computational complexity of each iteration, but also give a better approximation to the ℓ_1 norm of tensor multi-rank, which can lead to a better performance. We also provide an efficient algorithm for solving the resulting optimization problem. Finally, we apply the proposed method to the TC and the TRPCA problems, and provide some theoretical analysis on the performance guarantees.

In summary, our main contributions include:

- We propose a new definition of tensor Schatten-p norm with some desirable properties, e.g., unitary invariance, convexity and differentiability. When $0 < p < 1$, it is tighter than TNN to approximate the ℓ_1 norm of tensor multi-rank, which is beneficial to LRTR problems.
- We prove the tensor Schatten-p norm surrogate theorem, which helps us to transform a non-convex problem into many convex sub-problems[4], and we give an efficient algorithm to solve the transformed model. We also give a proof of the convergence of our algorithm. Our method can not only reduce the computational complexity of each iteration significantly when dealing with large-scale tensors, but also maintain the balance among factor tensors.
- We provide the sufficient conditions for exact recovery using a general linear operator and the error bounds based on some assumptions when there exists noise. For ensuring the performance of the TC problem, we give a theoretical analysis of exact completion.

We apply our proposed t-Schatten-p norm to the TRPCA and the TC problems. Experimental results on synthetic and real-world datasets verify the advantages of our method.

II. Notations and Preliminaries

In this section, we introduce some notations and necessary definitions which will be used later.

Tensors are represented by uppercase curlycue letters, e.g., $\mathcal{T}$. Matrices are represented by boldface uppercase letters, e.g., $\mathbf{M}$. Vectors are represented by boldface lowercase letters, e.g., $\mathbf{v}$. Scalars are represented by lowercase letters, e.g., c.

For a given 3-order tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we use $\mathbf{T}^{(k)}$ to represent its k-th frontal slice $\mathcal{T}(:,:,k)$. Its (i,j,k)-th entry is represented as $\mathcal{T}_{ijk}$. We use $\overline{\mathcal{T}}$ to represent the result of discrete Fourier transformation of $\mathcal{T}$ along the 3-rd dimension, corresponding to Matlab operator $\overline{\mathcal{T}} = \text{fft}(\mathcal{T}, [\,], 3)$. This also implies $\mathcal{T} = \text{ifft}(\overline{\mathcal{T}}, [\,], 3)$. $\overline{\mathbf{T}}^{(i)}$ denotes the i-th frontal slice of $\overline{\mathcal{T}}$. And the block circulant matrix associated to a 3-order tensor $\mathcal{T}$ is represented by $\text{bcirc}(\mathcal{T}) \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$:

$$\text{bcirc}(\mathcal{T}) = \begin{bmatrix} \mathbf{T}^{(1)} & \mathbf{T}^{(n_3)} & \cdots & \mathbf{T}^{(2)} \\ \mathbf{T}^{(2)} & \mathbf{T}^{(1)} & \cdots & \mathbf{T}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}^{(n_3)} & \mathbf{T}^{(n_3-1)} & \cdots & \mathbf{T}^{(1)} \end{bmatrix}.$$

As for block unfolding $\mathcal{T}$ and its inverse operation, we use the following operators:

$$\text{unfold}(\mathcal{T}) = \begin{bmatrix} \mathbf{T}^{(1)} \\ \mathbf{T}^{(2)} \\ \vdots \\ \mathbf{T}^{(n_3)} \end{bmatrix}, \quad \text{fold}(\text{unfold}(\mathcal{T})) = \mathcal{T}.$$

Then we define the t-product between two 3-order tensors as:

Definition 1: (t-product) [1] Let $\mathcal{A} \in \mathbb{R}^{n_1 \times s \times n_3}$, $\mathcal{B} \in \mathbb{R}^{s \times n_2 \times n_3}$. Then the t-product is defined as:

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})). \tag{5}$$

Here $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Note that if $n_3 = 1$, the operator $*$ reduces to matrix multiplication.

For tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, Kilmer *et al.* [1] point out that the block circulant matrix $\text{bcirc}(\mathcal{T})$ can be diagonalized by a specific matrix. We denote F_{n_3} as the $\mathcal{R}^{n_3 \times n_3}$ DFT matrix, and $F_{n_3}^H$ denotes the conjugate transpose of F_{n_3}. I_{n_1} and I_{n_2} are n_1-order and n_2-order identity matrices, respectively. Then using Kronecker product we have [1]:

$$(F_{n_3} \otimes I_{n_1}) \cdot \text{bcirc}(\mathcal{T}) \cdot (F_{n_3}^H \otimes I_{n_2}) = \begin{bmatrix} \overline{\mathbf{T}}^{(1)} & & & \\ & \overline{\mathbf{T}}^{(2)} & & \\ & & \ddots & \\ & & & \overline{\mathbf{T}}^{(n_3)} \end{bmatrix}.$$

Then the t-product can be calculated as follows:

Property 1: [1] Let $\mathcal{A} \in \mathbb{R}^{n_1 \times s \times n_3}$, $\mathcal{B} \in \mathbb{R}^{s \times n_2 \times n_3}$. Then the t-product is equivalent to matrix product of $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$:

$$\mathcal{T} = \mathcal{A} * \mathcal{B} \Longleftrightarrow \overline{\mathbf{T}}^{(k)} = \overline{\mathbf{A}}^{(k)} \overline{\mathbf{B}}^{(k)}, k = 1, \ldots, n_3. \tag{6}$$

Remark 1: In this paper, we use $\boxtimes$ to represent frontal-slice-wise matrix multiplication between tensors $\mathcal{A}$ and $\mathcal{B}$. Then

$$\mathcal{T} = \mathcal{A} * \mathcal{B} \Longleftrightarrow \overline{\mathbf{T}}^{(k)} = \overline{\mathbf{A}}^{(k)} \overline{\mathbf{B}}^{(k)} \Longleftrightarrow \overline{\mathcal{T}} = \overline{\mathcal{A}} \boxtimes \overline{\mathcal{B}}. \tag{7}$$

The relations of inner products (and Frobebius norm) in the time and the frequency domains are as follows.

Property 2: [3] Let $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then:

$$(1)\ \langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{n_3} \left\langle \overline{\mathcal{A}}, \overline{\mathcal{B}} \right\rangle,$$

$$(2)\ \|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \frac{1}{\sqrt{n_3}} \left\|\overline{\mathcal{A}}\right\|_F. \tag{8}$$

[4]But the whole optimization problem is still nonconvex due to the multilinear constraint.

For the definitions of tensor transpose $\mathcal{T}^*$ in Definition 8, Identity tensor $\mathcal{I}$ in Definition 9, and Orthogonal tensor in Definition 10, please refer to the Appendix. By using these notations, tensor Singular Value Decomposition (t-SVD) and Tensor Nuclear Norm (TNN) are defined as follows.

Definition 2: (t-SVD) [1] Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then there exist $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ such that:

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*, \tag{9}$$

where $\mathcal{U} * \mathcal{U}^* = \mathcal{I}$, $\mathcal{V} * \mathcal{V}^* = \mathcal{I}$, and $\mathcal{S}$ is a frontal-slice-diagonal tensor.

Definition 3: (TNN) [3] The tensor nuclear norm of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{T}\|_*$, is defined as the average of the nuclear norm of all the frontal slices of $\overline{\mathcal{T}}$ as follow:

$$\|\mathcal{T}\|_* := \frac{1}{n_3} \sum_{i=1}^{n_3} \|\overline{\mathbf{T}}^{(i)}\|_*. \tag{10}$$

Furthermore, the tensor spectral norm, tensor multi-rank and tubal-rank are defined by using t-SVD as follows:

Definition 4: (Tensor spectral norm) [3] Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The tensor spectral norm of $\mathcal{T}$ is defined as $\|\mathcal{T}\| := \max_i \{\|\overline{\mathbf{T}}^{(i)}\|\}$.

By using the Von Neumann's inequality, it is easy to prove that the dual norm of tensor spectral norm is the tensor nuclear norm and vice versa.

Definition 5: (Tensor multi-rank and tubal-rank) [4] Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the tensor multi-rank of $\mathcal{T}$ is a vector $\mathbf{r} \in \mathbb{R}^{n_3}$ with its i-th entry as the rank of the i-th frontal slice of $\overline{\mathcal{T}}$, i.e., $r_i = \text{rank}(\overline{\mathbf{T}}^{(i)})$. The tensor tubal-rank of $\mathcal{T}$, denoted as $\text{rank}_t(\mathcal{T})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ is from the t-SVD of $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$.

III. Main Result

Comparing with the relations between matrix nuclear norm and matrix schatten-p norm, we propose a new definition of tensor Schatten-p norm (t-Schatten-p norm) based on TNN and t-SVD as:

Definition 6: (Tensor Schatten-p Norm) Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and its tensor singular value decomposition be $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$. Then the tensor schatten-p norm is defined as:

$$\|\mathcal{T}\|_{S_p} := \left(\frac{1}{n_3} \sum_{i=1}^{n_3} \|\overline{\mathbf{T}}^{(i)}\|_{S_p}^p \right)^{\frac{1}{p}},$$

$$\text{i.e.,} \quad \|\mathcal{T}\|_{S_p} := \left(\frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{i=1}^{\min\{n_1,n_2\}} \left|\overline{\mathcal{S}}_{iik}\right|^p \right)^{\frac{1}{p}}. \tag{11}$$

When $p = 1$ it becomes the tensor nuclear norm, which is similar to the matrix case.

Obviously, t-Schatten-p norm satisfies: (1) $\|\mathcal{T}\|_{S_p} \geq 0$ with equality holding if and only if $\mathcal{T}$ is zero; (2) $\|\alpha \mathcal{T}\|_{S_p} = \alpha \|\mathcal{T}\|_{S_p}$.

A. Algebraic Properties

Our proposed t-Schatten-p norm has some properties similar to the matrix Schatten-p norm. The followings are some of the properties of $\|\mathcal{T}\|_{S_p}$ and $\|\mathcal{T}\|_{S_p}^p$ that we use in this paper. For the proofs, please refer to the Supplementary Materials.

Proposition 1: (Unitary Invariance) Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ be orthogonal tensors. The tensor Schatten-p norm defined in Eq. (11) is unitary invariant, i.e.,

$$\|\mathcal{T}\|_{S_p} = \|\mathcal{U} * \mathcal{T}\|_{S_p} = \|\mathcal{T} * \mathcal{V}^*\|_{S_p} = \|\mathcal{U} * \mathcal{T} * \mathcal{V}^*\|_{S_p}. \tag{12}$$

Proposition 2: Given a 3-order tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, when $p \geq 1$, $\|\mathcal{T}\|_{S_p}^p$ is convex w.r.t. $\mathcal{T}$. In other words, it satisfies the inequality for any $\lambda \in (0, 1)$:

$$\|\lambda \mathcal{A} + (1-\lambda)\mathcal{B}\|_{S_p}^p \leq \lambda \|\mathcal{A}\|_{S_p}^p + (1-\lambda) \|\mathcal{B}\|_{S_p}^p, \quad p \geq 1. \tag{13}$$

And when $0 < p < 1$, $\|\mathcal{T}\|_{S_p}^p$ is non-convex w.r.t. $\mathcal{T}$.

With this proposition, when $p \geq 1$, $\|\mathcal{T}\|_{S_p}^p$ can be applied to many convex optimization problems. For a certain convex function, whether it is differentiable or not is very important. The next proposition gives the answer when $p \geq 1$.

Proposition 3: Given a 3-order tensor $\mathcal{T}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and that the skinny t-SVD of $\mathcal{T}_0$ is $\mathcal{U} * \mathcal{S} * \mathcal{V}^*$. When $p > 1$, the gradient of $\|\mathcal{T}\|_{S_p}^p$ at $\mathcal{T}_0$ has the following form:

$$\nabla_{\mathcal{T}_0} \|\mathcal{T}\|_{S_p}^p = p\, \mathcal{U} * \mathcal{D} * \mathcal{V}^*, \quad \text{with } \overline{\mathcal{D}} = \overline{\mathcal{S}}^{p-1}, \ p > 1. \tag{14}$$

Moreover, when $p = 1$, the subdifferential of $\|\mathcal{T}\|_{S_p}^p$ at $\mathcal{T}_0$ is:

$$\partial_{\mathcal{T}_0} \|\mathcal{T}\|_{S_p}^p = \{\mathcal{U} * \mathcal{V}^* + \mathcal{W} | \mathcal{U}^* * \mathcal{W} = \mathcal{O}, \mathcal{W} * \mathcal{V} = \mathcal{O}\}. \tag{15}$$

For models which need to minimze $\|\mathcal{T}\|_{S_p}^p$, Proposition 3 indicates that when $p \geq 1$ we can use gradient-based methods to solve them.

B. Unified Surrogate Theorem

For the matrix case, there exist many surrogates for a specific matrix Schatten-p norm, such as $p = 1, 2/3, 1/2$ [26]–[28]. Xu *et al.* [25] give a general result of unified surrogates for the matrix Schatten-p norm.

Lemma 1: (Multi-Schatten-p Norm Surrogate) [25] Given I $(I \geq 2)$ matrices $\mathbf{X}_i$ $(i = 1, \ldots, I)$, where $\mathbf{X}_1 \in \mathbb{R}^{m \times d_1}$, $\mathbf{X}_i \in \mathbb{R}^{d_{i-1} \times d_i}$ $(i = 2, \ldots, I-1)$, $\mathbf{X}_I \in \mathbb{R}^{d_{I-1} \times n}$, and $\mathbf{X} \in \mathbb{R}^{m \times n}$ with $rank(\mathbf{X}) = r \leq \min\{d_i, i = 1, \ldots, I-1\}$, for any $p, p_1, \ldots, p_I > 0$ satisfying $1/p = \sum_{i=1}^I 1/p_i$, we have

$$\frac{1}{p} \|\mathbf{X}\|_{S_p}^p = \min_{\mathbf{X}_i : \mathbf{X} = \prod_{i=1}^I \mathbf{X}_i} \sum_{i=1}^I \frac{1}{p_i} \|\mathbf{X}_i\|_{S_{p_i}}^{p_i}. \tag{16}$$

This lemma indicates that for any given Schatten-p norm, we can get the same value by solving a minimization problem. The following Theorem 1 points out that for our proposed t-Schatten-p norm, this rule holds too.

Theorem 1: (Multi-Tensor-Schatten-p Norm Surrogate) Given $I (I \geq 2)$ tensors $\mathcal{T}_i$ $(i = 1, \ldots, I)$, where $\mathcal{T}_1 \in \mathbb{R}^{m \times d_1 \times k}$, $\mathcal{T}_i \in \mathbb{R}^{d_{i-1} \times d_i \times k}$ $(i = 2, \ldots, I-1)$, $\mathcal{T}_I \in \mathbb{R}^{d_{I-1} \times n \times k}$, and $\mathcal{T} \in$

$\mathbb{R}^{m\times n\times k}$ with $\text{rank}_t(\mathcal{T}) = r \leq \min\{d_i, i = 1, \ldots, I-1\}$, for any $p, p_1, \ldots, p_I > 0$ satisfying $1/p = \sum_{i=1}^{I} 1/p_i$, we have

$$\frac{1}{p}\|\mathcal{T}\|_{S_p}^p = \min_{\mathcal{T}_i:\mathcal{T}=\mathcal{T}_1*\cdots*\mathcal{T}_I} \sum_{i=1}^{I} \frac{1}{p_i}\|\mathcal{T}_i\|_{S_{p_i}}^{p_i}. \tag{17}$$

In the previous section, we mention that when $p \geq 1$ the t-Schatten-p norm $\|\mathcal{T}\|_{S_p}^p$ is convex. When $p < 1$, we can use Theorem 1 to convert the non-convex function $\|\mathcal{T}\|_{S_p}^p$ into weighted sum of convex functions $\|\mathcal{T}_i\|_{S_{p_i}}^{p_i}$ with $p_i \geq 1$ and $\mathcal{T} = \mathcal{T}_1 * \cdots * \mathcal{T}_I$. Note that p_i can also be less than 1, i.e., $0 < p_i < 1$.

C. Tightness

Lu *et al.* [3] point out that if the tensor average rank is defined as $\text{rank}_a(\mathcal{T}) = \frac{1}{n_3}\sum_{i=1}^{n_3} \text{rank}(\overline{\mathbf{T}}^{(i)})$, then TNN is the convex envelope of the tensor average rank within the unit ball of the tensor spectral norm. Here we claim that when $0 < p < 1$, our $\|\mathcal{T}\|_{S_p}^p$ is a tighter non-convex envelope of the tensor average rank within the same unit ball.

Proposition 4: For a given 3-order tensor $\mathcal{T}$ and $0 < p \leq 1$, $\|\mathcal{T}\|_{S_p}^p$ is an envelope of the tensor average rank within the unit ball of the tensor spectral norm. If we set $p = 1$, it becomes TNN, which is the tightest convex envelope. Otherwise, it is a non-convex envelope of the tensor multi-rank, which is tighter than TNN in the sense of $\|\mathcal{T}\|_* \leq \|\mathcal{T}\|_{S_p}^p \leq \frac{1}{n_3}\|\text{rank}_m(\mathcal{T})\|_1$.

Zhang *et al.* [4] give another definition of TNN and prove that it is the tightest convex envelope to ℓ_1 norm of the tensor multi-rank within a unit ball. It is easy to find that these two conclusions are equivalent.

Considering the LRTR problem in Eq. (1), we usually need to use a relaxed function to replace the non-smooth rank function. By using Proposition 4, we transform the recovery problem into the following:

$$\begin{aligned}&\min_{\mathcal{X}} \|\mathcal{X}\|_{S_p}^p,\\ &\text{s.t. } \Psi(\mathcal{X}) = \mathcal{T},\end{aligned} \tag{18}$$

where $\mathcal{T}$ is the observed measurement through a linear operator $\Psi(\cdot)$. We aim to recover the clean tensor $\mathcal{X}$ based on the observed corrupted/missing tensor $\mathcal{T}$.

D. Advantages and Disadvantages

The main advantage is that our $\|\mathcal{T}\|_{S_p}^p$ is a tighter non-convex envelope of the tensor average rank within the unit ball of the tensor spectral norm, as shown in Proposition 4. Liu *et al.* [29] point out that, in matrix case, when using the matrix factorization or Schatten-p norm $(p = 2/3)$ the recovery condition is weaker than the convex optimization based on nuclear norm. That is to say, when using a better approximation, we may get a better solution. Our experiments in Section VI also verify this statement.

However, the disadvantage is also obvious. Compared with TNN in [4], $\|\mathcal{T}\|_{S_p}^p$ is non-convex and non-smooth when $0 < p < 1$. It would be hard to obtain a strong performance guarantee as done in the convex programs, e.g., [22]. Even in the matrix case, it is still unknown under what conditions a specific optimization procedure for Schatten-p norm can produce an optimal solution that exactly recovers the target matrix. The same is true for the tensor case.

IV. Application and Optimization

In this section, we propose a general LRTR model based on the t-Schatten-p norm and give a feasible algorithm to solve it.

A. Model

In practical applications, the observed tensor $\mathcal{T}$ is inevitably contaminated by noise. Therefore we add a noise tensor and a noise regularization to the model in Eq. (18):

$$\begin{aligned}&\min_{\mathcal{X},\mathcal{E}} \|\mathcal{X}\|_{S_p}^p + \lambda g(\mathcal{E}),\\ &\text{s.t. } \Psi(\mathcal{X}) + \mathcal{E} = \mathcal{T},\end{aligned} \tag{19}$$

where $\mathcal{T}$ is the observed tensor, $\mathcal{E}$ is a noise tensor and $g(\cdot)$ denotes the noise regularization. In specific problems, if we assume that the noise follows the Gaussian distribution or the Laplacian distribution, $g(\mathcal{E})$ can be chosen as $\|\mathcal{E}\|_F^2$ or $\|\mathcal{E}\|_1$, respectively.

When $p \geq 1$, the problem in Eq. (19) can be solved by many convex optimization methods directly. If $0 < p < 1$, the t-Schatten-p norm becomes non-convex. Then we can use Theorem 1 to convert the non-convex function into sum of several convex functions. The following proposition provides the guarantee of this idea.

Proposition 5: [25] For any $0 < p < 1$, there always exist $I \in N$ and p_i such that $1/p = \sum_{i=1}^{I} 1/p_i$, where all p_i satisfy one of the cases: (a) $p_i \geq 1$ or (b) $p_i > 1$.

Given $I\ (I \geq 2)$ and $i = 1, \ldots, I$, for any $p, p_i > 0$ satisfying $1/p = \sum_{i=1}^{I} 1/p_i$, we assume that $\mathcal{X} = \mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I$, then (19) can be converted to:

$$\begin{aligned}&\min_{\{\mathcal{X}_i\},\mathcal{E}} \sum_{i=1}^{I} \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i} + \lambda g(\mathcal{E}),\\ &\text{s.t. } \Psi(\mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I) + \mathcal{E} = \mathcal{T}.\end{aligned} \tag{20}$$

If $p_i \geq 1$ holds for all i, the optimization of problem in Eq. (20) becomes multi-convex. Thus if we apply the block coordinate descent method to solve Eq. (20), each $\mathcal{X}_i$ can be efficiently updated by convex optimization.

B. Optimization

Different from the matrix case in [26], we need to introduce an intermediate tensor $\mathcal{G}$ to separate $\{\mathcal{X}_i\}$ from $\Psi(\cdot)$. If not, calculating the sub-gradient of $\|\Psi(\mathcal{X}_1 * \mathcal{X}_2)\|_F$ may be a difficult problem for certain Ψ, such as $\mathcal{P}_\Omega$ in the TC problem. Then by adding an equality constraint, Eq. (20) can be rewritten as:

$$\begin{aligned}&\min_{\{\mathcal{X}_i\},\mathcal{E}} \sum_{i=1}^{I} \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i} + \lambda g(\mathcal{E}),\\ &\text{s.t. } \Psi(\mathcal{G}) + \mathcal{E} = \mathcal{T},\\ &\quad\ \ \mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I = \mathcal{G}.\end{aligned} \tag{21}$$

Here we solve Eq. (21) by a new method based on LADMPSAP [30] and BCD [31]. By introducing Lagrange multipliers $\mathcal{Y}$ and $\mathcal{Z}$, the augmented Lagrangian function of (21) is given as follows:

$$
\begin{aligned}
\mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) &= \sum_{i=1}^{I} \frac{1}{p_i} \|\mathcal{X}_i\|_{S_{p_i}}^{p_i} + \lambda g(\mathcal{E}) \\
&+ \langle \mathcal{Y}, \Psi(\mathcal{G}) + \mathcal{E} - \mathcal{T} \rangle + \frac{\rho_1}{2} \|\Psi(\mathcal{G}) + \mathcal{E} - \mathcal{T}\|_F^2 \\
&+ \langle \mathcal{Z}, \mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I - \mathcal{G} \rangle \\
&+ \frac{\rho_2}{2} \|\mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I - \mathcal{G}\|_F^2 . \quad (22)
\end{aligned}
$$

All $\{\mathcal{X}_i\}$ in (22) need to be updated sequentially. Note that different order of updating $\mathcal{X}_i$ may lead to different convergence rates. We update $\mathcal{X}_1$ and $\mathcal{X}_I$ first, and then update others in proper order.

1) Update $\mathcal{X}_i$: Assume that we have already updated $\mathcal{X}_1^k, \mathcal{X}_2^k, \ldots, \mathcal{X}_{i-1}^k$. Let $\mathcal{Q}_{1:(i-1)} = \mathcal{X}_1^k * \cdots * \mathcal{X}_{i-1}^k$ and $\mathcal{Q}_{(i+1):I} = \mathcal{X}_{i+1}^{k-1} * \cdots * \mathcal{X}_I^{k-1}$. Then the sub-problem for updating $\mathcal{X}_i^k$ can be written as:

$$
\begin{aligned}
\mathcal{X}_i^k = \arg\min_{\mathcal{X}_i} & \frac{1}{p_i} \|\mathcal{X}_i\|_{S_{p_i}}^{p_i} \\
&+ \frac{\rho_2}{2} \|\mathcal{Q}_{1:(i-1)} * \mathcal{X}_i * \mathcal{Q}_{(i+1):I} - \mathcal{G}^{k-1} + \mathcal{Z}^{k-1}/\rho_2\|_F^2 .
\end{aligned} \quad (23)
$$

Calculating (23) directly is difficult. Letting $f_i^k(\mathcal{X}_i) = \frac{1}{2} \|\mathcal{Q}_{1:(i-1)} * \mathcal{X}_i * \mathcal{Q}_{(i+1):I} - \mathcal{G}^{k-1} + \mathcal{Z}^{k-1}/\rho_2\|_F^2$ and further linearizing $f_i^k(\mathcal{X}_i)$ at point $\mathcal{X}_i^{k-1}$, the sub-problem becomes:

$$
\begin{aligned}
\mathcal{X}_i^k &= \arg\min_{\mathcal{X}_i} \frac{1}{p_i} \|\mathcal{X}_i\|_{S_{p_i}}^{p_i} + \rho_2 \Bigg(\langle \nabla f_i^k(\mathcal{X}_i^{k-1}), \mathcal{X}_i - \mathcal{X}_i^{k-1} \rangle \\
&\qquad + \frac{L_i^{k-1}}{2} \|\mathcal{X}_i - \mathcal{X}_i^{k-1}\|_F^2 \Bigg) \\
&= \arg\min_{\mathcal{X}_i} \frac{1}{p_i} \|\mathcal{X}_i\|_{S_{p_i}}^{p_i} \\
&\qquad + \frac{\rho_2 L_i^{k-1}}{2} \left\| \mathcal{X}_i - \mathcal{X}_i^{k-1} + \frac{\nabla_i^k f(\mathcal{X}_i^{k-1})}{L_i^{k-1}} \right\|_F^2 , \quad (24)
\end{aligned}
$$

where $\nabla f_i^k(\mathcal{X}_i^{k-1})$ is the gradient of $f_i^k(\mathcal{X}_i)$ at $\mathcal{X}_i^{k-1}$:

$$
\begin{aligned}
&\nabla f_i^k(\mathcal{X}_i^{k-1}) = \mathcal{Q}_{1:(i-1)}^* \\
&* \left(\mathcal{Q}_{1:(i-1)} * \mathcal{X}_i^{k-1} * \mathcal{Q}_{(i+1):I} - \mathcal{G}^{k-1} + \mathcal{Z}^{k-1}/\rho_2 \right) * \mathcal{Q}_{(i+1):I}^* .
\end{aligned} \quad (25)
$$

L_i^{k-1} is the Lipschitz constant of $\nabla f_i^k(\mathcal{X}_i^{k-1})$, and it can be chosen as the tensor spectral norm of $\nabla f_i^k(\mathcal{X}_i^{k-1})$. The right hand side of Eq. (24) is the proximal mapping of $\mathcal{X}_i$ and can be solved by off-the-shelf algorithms:

$$
\mathcal{X}_i^k = \mathbf{Prox}_{p_i \rho_2 L_i^{k-1}, \|\cdot\|_{S_{p_i}}^{p_i}} \left(\mathcal{X}_i^{k-1} - \frac{\nabla_i^k f(\mathcal{X}_i^{k-1})}{L_i^{k-1}} \right), \quad (26)
$$

where $\mathbf{Prox}_{\lambda, f}(z) := \arg\min_x f(x) + \frac{\lambda}{2}\|x - z\|_F^2$.

There are some special cases of p_i that have closed-form solutions. When $p_i = 1$, the problem in Eq. (26) can be solved by the tensor Singular Value Thresholding (tSVT) operator [4]. When $p_i = 2$, the problem in Eq. (26) can be solved by calculating the unique critical point of quadratic function, or using Theorem 3.1 in [13] by setting $\gamma = 0$. When $p_i < 1$, the problem in Eq. (26) becomes non-smooth and non-convex. We can use the generalized iterated shrinkage algorithm (GISA) [32] to obtain a high-precision solution.

2) Update $\mathcal{G}$: By fixing other variables, we have the following sub-problem to update $\mathcal{G}$:

$$
\begin{aligned}
\mathcal{G}^k = \arg\min_{\mathcal{G}} & \frac{\rho_1}{2} \left\| \Psi(\mathcal{G}) + \mathcal{E}^{k-1} - \mathcal{T} + \mathcal{Y}^{k-1}/\rho_1 \right\|_F^2 \\
&+ \frac{\rho_2}{2} \left\| \mathcal{X}_1^k * \mathcal{X}_2^k * \cdots * \mathcal{X}_I^k - \mathcal{G} + \mathcal{Z}^{k-1}/\rho_2 \right\|_F^2 . \quad (27)
\end{aligned}
$$

Different linear operators Ψ result in different solutions of Eq. (27). For example, in TRPCA problems Ψ is an identity operator and $\mathcal{G}^k$ is given by:

$$
\mathcal{G}^k = \frac{\rho_1 \left(\mathcal{T} - \mathcal{E}^{k-1} - \mathcal{Y}^{k-1}/\rho_1 \right) + \rho_2 \left(\mathcal{Q}_{1:I}^k + \mathcal{Z}^{k-1}/\rho_2 \right)}{\rho_1 + \rho_2}, \quad (28)
$$

where $\mathcal{Q}_{1:I}^k = \mathcal{X}_1^k * \mathcal{X}_2^k * \cdots * \mathcal{X}_I^k$. In TC problems Ψ is a projection operator $\mathcal{P}_\Omega$. If we devide the identity operator into $\mathcal{P}_\Omega$ and $\mathcal{P}_{\bar{\Omega}}$, then $\mathcal{G}^k$ is given by:

$$
\begin{aligned}
&\mathcal{G}^k = \mathcal{P}_{\bar{\Omega}} \left(\mathcal{Q}_{1:I}^k + \mathcal{Z}^{k-1}/\rho_2 \right) \\
&+ \mathcal{P}_\Omega \left(\frac{\rho_1 \left(\mathcal{T} - \mathcal{E}^{k-1} - \mathcal{Y}^{k-1}/\rho_1 \right) + \rho_2 \left(\mathcal{Q}_{1:I}^k + \mathcal{Z}^{k-1}/\rho_2 \right)}{\rho_1 + \rho_2} \right).
\end{aligned} \quad (29)
$$

3) Update $\mathcal{E}$: By fixing other variables, we have the following sub-problem to update $\mathcal{E}$:

$$
\begin{aligned}
&\mathcal{E}^k = \arg\min_{\mathcal{E}} \lambda g(\mathcal{E}) + \frac{\rho_1}{2} \|\Psi(\mathcal{G}^k) + \mathcal{E} - \mathcal{T} + \mathcal{Y}^{k-1}/\rho_1\|_F^2, \\
&\text{i.e., } \mathcal{E}^k = \mathbf{Prox}_{\rho_1/\lambda, g(\cdot)} \left(\mathcal{T} - \Psi\left(\mathcal{G}^k\right) - \mathcal{Y}^{k-1}/\rho_1 \right).
\end{aligned} \quad (30)
$$

The problem in Eq. (30) usually has a closed-form solution for a specific $g(\cdot)$. If $g(\cdot)$ is chosen as $\|\mathcal{E}\|_F^2$, $\mathcal{E}^k$ is given by:

$$
\mathcal{E}^k = \frac{\mathcal{T} - \Psi\left(\mathcal{G}^k\right) - \mathcal{Y}^{k-1}/\rho_1}{1 + \rho_1/\lambda}. \quad (31)
$$

If $g(\cdot)$ is chosen as $\|\mathcal{E}\|_1$, and let $\mathcal{R} = \mathcal{T} - \Psi\left(\mathcal{G}^k\right) - \mathcal{Y}^{k-1}/\rho_1$, then $\mathcal{E}^k$ is given by:

$$
\mathcal{E}^k = \operatorname{sign}(\mathcal{R}) \max\left\{ |\mathcal{R}| - \frac{\lambda}{\rho_1}, 0 \right\}. \quad (32)
$$

*4) Update $\mathcal{Y}$ **and** $\mathcal{Z}$:* After updating $\{\mathcal{X}_i^k\}, \mathcal{G}^k$ and $\mathcal{E}^k$, the Lagrange multipliers $\mathcal{Y}$ and $\mathcal{Z}$ are updated by:

$$
\begin{cases}
\mathcal{Y}^k := \mathcal{Y}^{k-1} + \rho_1 \left(\Psi(\mathcal{G}^k) + \mathcal{E}^k - \mathcal{T} \right), \\
\mathcal{Z}^k := \mathcal{Z}^{k-1} + \rho_2 \left(\mathcal{X}_1^k * \mathcal{X}_2^k * \cdots * \mathcal{X}_I^k - \mathcal{G}^k \right).
\end{cases} \quad (33)
$$

For the penalty parameters ρ_1 and ρ_2, Lin *et al.* [30] further suggest increasing them gradually. We summarize the algorithm for solving the problem in Eq. (21) in Algorithm 1.

Algorithm 1: Solving Problem (21).

Input: The observed tensor data $\mathcal{T}$ and parameters $\lambda, \rho_1^0, \rho_2^0, \rho_{1,max}, \rho_{2,max}. \eta, \varepsilon$.
Initialize: $\left\{\mathcal{X}_i^0\right\}, \mathcal{G}^0, \mathcal{E}^0, \mathcal{Y}^0, \mathcal{Z}^0$.
While not converge **do**
1) Update $\mathcal{X}_i^k$ by solving (26) for $i = 1, \ldots, I$.
2) Update $\mathcal{G}^k$ by solving (27).
3) Update $\mathcal{E}^k$ by solving (30).
4) Update $\mathcal{Y}^k$ and $\mathcal{Z}^k$ by (33).
5) Update ρ_1^k and ρ_2^k by $\begin{cases}\rho_1^k = \min\{\eta\rho_1^{k-1}, \rho_{1,max}\} \\ \rho_2^k = \min\{\eta\rho_2^{k-1}, \rho_{2,max}\}\end{cases}$
6) Check the convergence condition: $\|\mathcal{X}_i^k - \mathcal{X}_i^{k-1}\| \leq \varepsilon_2$, $\|\mathcal{G}^k - \mathcal{G}^{k-1}\| \leq \varepsilon_2$, and $\|\mathcal{E}^k - \mathcal{E}^{k-1}\| \leq \varepsilon_2$.
7) $k \leftarrow k + 1$.
end While
Output: The factor tensors $\{\mathcal{X}_i\}$, the noise tensor $\mathcal{E}$ and the intermediate tensor $\mathcal{G}$.

C. Convergence Analysis

In general, it is hard to provide the convergence for the ADMM based method with a Burer-Monteiro factorization constraint. Fortunately, due to the separability of the proposed objective, we can still prove the convergence of our algorithm by assuming the smoothness of the noise regularization $g(\cdot)$ in Eq. (21) and all $p_i \geq 1$. Note that in the following Theorem 2, p can be chosen in the range $(0, +\infty)$ as long as $1/p = \sum_{i=1}^I 1/p_i$ holds.

Theorem 2: If the optimization problem in Eq. (21) satisfies the following conditions: (a) the function $g(\cdot)$ in Eq. (21) is smooth, convex, and coercive; (b) $p_i \geq 1$ for $i = 1, \ldots, I$; (c) ρ_1 and ρ_2 in Eq. (22) are sufficiently large, then the sequence $\{\mathcal{X}_i^k, \mathcal{G}^k, \mathcal{E}^k, \mathcal{Y}^k, \mathcal{Z}^k\}$ generated in Algorithm 1 satisfies the following properties:

(1) The augmented Lagrangian function (22) is monotonically decreasing, i.e. for some $c > 0$,

$$\begin{aligned}&\mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}^+, \mathcal{Z}^+)\\ &\geq c(\|(\mathcal{X}_i - \mathcal{X}_i^+)\|_F^2 + \|\mathcal{G} - \mathcal{G}^+\|_F^2 + \|\mathcal{E} - \mathcal{E}^+\|_F^2);\end{aligned}$$

(2) $\|\mathcal{X}_i^+ - \mathcal{X}_i\| \to 0, \|\mathcal{G}^+ - \mathcal{G}\| \to 0, \|\mathcal{E}^+ - \mathcal{E}\| \to 0$;

(3) The sequence $\{\mathcal{X}_i^k, \mathcal{G}^k, \mathcal{E}^k, \mathcal{Y}^k, \mathcal{Z}^k\}$ are bounded;

(4) Any accumulation point of the sequence $\{\mathcal{X}_i^k, \mathcal{G}^k, \mathcal{E}^k, \mathcal{Y}^k, \mathcal{Z}^k\}$ is a constrained stationary point.

D. Complexity Analysis

For Algorithm 1, different p_i's result in different complexities. Here we choose the widely used case $p_1 = p_2 = 1$ to analyze. If $\mathcal{X}_1 \in \mathbb{R}^{n_1 \times d \times n_3}$, $\mathcal{X}_2 \in \mathbb{R}^{d \times n_2 \times n_3}$ and $\text{rank}_t(\mathcal{X}) = r\ (r \leq d)$, then the per-iteration complexity in Algorithm 1 is $\mathcal{O}\left((n_1 + n_2)n_3 d \log n_3 + (n_1 + n_2)n_3 d^2 + n_1 n_2 n_3 d\right)$. One iteration means updating all variables once in order. As for TNN in [3], the computational complexity at each iteration is $\mathcal{O}(n_1 n_2 n_3(\log n_3 + \min\{n_1, n_2\}))$. Obviously, when $d \ll \min\{n_1, n_2\}$, our method is much more efficient than TNN based methods in each iteration. Due to the convexity of TNN, the related problem usually needs fewer iterations to converge. But when the tensor rank is large or the noise is great, it may perform worse than our proposed methods. Our experiments in Section VI.B also verify this conclusion.

V. Recovery Guarantees

In this section, we provide theoretical guarantees for LRTR problems based on our proposed t-Schatten-p norm, which aim to recover low-rank tensors from linear observations. For the proofs of our theorems, please refer to the Supplementary Materials.

A. Null Space Property (NSP)

NSP is widely used in the theoretical analysis of recovering sparse vectors and low-rank matrices [33], [34]. Here we give a sufficient condition for exactly recovering the low-rank tensor $\widehat{\mathcal{X}}$ in Eq. (18) by the following model:

$$\begin{aligned}\min_{\{\mathcal{X}_i\}}\ & \sum_{i=1}^I \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i},\\ \text{s.t.}\ & \Psi(\mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I) = \mathcal{T}.\end{aligned} \tag{34}$$

Assume $\widehat{\mathcal{X}} = \widehat{\mathcal{U}} * \widehat{\mathcal{S}} * \widehat{\mathcal{V}}^*$ to be the true tensor in Eq. (18) with $\text{rank}_t(\widehat{\mathcal{X}}) = r$, and $\widehat{\mathcal{X}} = \widehat{\mathcal{X}}_1 * \cdots * \widehat{\mathcal{X}}_I$ with $\widehat{\mathcal{X}}_1 = \widehat{\mathcal{U}} * \widehat{\mathcal{S}}^{p/p_1}$, $\widehat{\mathcal{X}}_2 = \widehat{\mathcal{S}}^{p/p_2}$, ..., and $\widehat{\mathcal{X}}_I = \widehat{\mathcal{S}}^{p/p_I} * \widehat{\mathcal{V}}^*$. $\mathcal{N}(\Psi) := \{\mathcal{X} : \Psi(\mathcal{X}) = \mathbf{0}\}$ denotes the null space of the linear operator Ψ. Then we have the following theorem:

Theorem 3: Assume $\widehat{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ to be the true tensor for Eq. (18) with tubal-rank $\text{rank}_t(\widehat{\mathcal{X}}) = r$, and $p \in (0, 1]$ with $\frac{1}{p} = \sum_{i=1}^I \frac{1}{p_i}$. In addition, for any $\mathcal{X} \in \{\mathcal{X}_i\}_{i=1}^I$ with $\mathcal{X} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$, $\min\{k_1, k_2\} \geq r$ holds. Then $\widehat{\mathcal{X}}$ is the unique optimal solution of Eq. (18) and can be uniquely recovered by Eq. (34), if for any $\mathcal{Z} = (\widehat{\mathcal{X}}_1 + \mathcal{W}_1) * \cdots * (\widehat{\mathcal{X}}_I + \mathcal{W}_I) - (\widehat{\mathcal{X}}_1 * \cdots * \widehat{\mathcal{X}}_I) \in \mathcal{N}(\Psi) \setminus \mathcal{O}$, where $\{\mathcal{W}_i\}$ have compatible dimensions and $\mathcal{N}(\Psi)$ denotes the null space of the linear operator Ψ, we have

$$\sum_{i=1}^{r}\sum_{j=1}^{n_3} \sigma_i^p\left(\overline{\mathbf{Z}}^{(j)}\right) < \sum_{i=r+1}^{\min\{n_1,n_2\}}\sum_{j=1}^{n_3} \sigma_i^p\left(\overline{\mathbf{Z}}^{(j)}\right). \tag{35}$$

Note that this condition is usually hard to be satisfied. Therefore, for specific problems we need to give some error bounds between the true tensors and the solutions by our algorithm.

B. Error Bound Analysis for Robust Tensor Recovery

In this section, we first introduce the following assumption of the general linear operator $\mathcal{A}$. Based on this assumption, we then give a theoretical analysis of the error bound for robust tensor recovery.

Assumption 1: [35] Suppose that there is a positive constant $\kappa(\mathcal{A})$ such that for $\Delta \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\Delta \in \mathcal{C}$, the general linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^l$ satisfies the following inequality:

$$\|\mathcal{A}(\Delta)\|_2 \geq \kappa(\mathcal{A})\|\Delta\|_F, \quad \Delta \in \mathcal{C}. \tag{36}$$

When $\mathcal{C}$ denotes the whole space $\mathbb{R}^{n_1 \times n_2 \times n_3}$, $\kappa(\mathcal{A})$ actually can be chosen as the smallest singular value of operator $\mathcal{A}$.

Based on Assumption 1, we provide the error bound for robust tensor recovery via Eqs. (19) and (20), which have noisy measurements.

Theorem 4: Assume that $\widehat{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a true tensor which satisfies the corrupted measurements $\Psi(\widehat{\mathcal{X}}) + \mathcal{E} = \mathcal{T}$, where $\mathcal{E}$ is the noise with $\|\mathcal{E}\|_F \leq \epsilon$. Let $(\widehat{\mathcal{X}}_1, \ldots, \widehat{\mathcal{X}}_I)$ be a critical point of Eq. (34) with the squared loss $\frac{\lambda}{2}\|\cdot\|_F^2$ and all $p_i \geq 1$. Here $\mathrm{rank}_t(\widehat{\mathcal{X}}) = r\ (r \leq d)$ and $d = \min\{\min\{p, q\} : \widehat{\mathcal{X}}_i \in \mathbb{R}^{p \times q \times l}, i = 1, \ldots, I\}$. If the linear operator Ψ satisfies the condition of Assumption 1 with a positive constant $\kappa(\Psi)$ on $\mathcal{R}^{n_1 \times n_2 \times n_3}$, then

$$\frac{\left\|\widehat{\mathcal{X}} - \widehat{\mathcal{X}}_1 * \cdots * \widehat{\mathcal{X}}_I\right\|_F}{\sqrt{n_1 n_2 n_3}} \leq \frac{\epsilon}{\kappa(\Psi)\sqrt{n_1 n_2 n_3}} + \frac{\sqrt{t}}{\lambda C_1 \kappa(\Psi)\sqrt{n_1 n_2 n_3}}, \tag{37}$$

where $t \geq d$ and C_1 is a constant related to $\{\widehat{\mathcal{X}}_i\}$. We give a lower bound for C_1 in the Supplementary Materials.

Theorem 4 claims that if Ψ satisfies the condition in Assumption 1, then there is an upper bound of the error between any critical point of Eq. (34) with the squared loss $\frac{\lambda}{2}\|\cdot\|_F^2$ and the true tensor in Eq. (19). The right hand side gives a rough guarantee for our proposed model. When the noise is small, the exact solution is close to the critical point.

C. Guarantee for Tensor Completion

The TC problem plays an important role in practical applications. However, the projection operator $\mathcal{P}_\Omega$ in Eq. (38) usually does not satisfy the RIP condition or Assumption 1 [13], so the TC problem should be treated as a special case. By setting Ψ as the projection operator $\mathcal{P}_\Omega$ in Eq. (34), we get the following formulation:

$$\begin{aligned} \min_{\{\mathcal{X}_i\}} \quad & \sum_{i=1}^{I} \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i}, \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I) = \mathcal{P}_\Omega(\mathcal{T}). \end{aligned} \tag{38}$$

Note that the error bound introduced in Theorem 4 usually does not hold. By using Theorem 8 in [35] we can deduce that the error bounds for the TC problem are related to $|\Omega|$ and rank of each frontal slice, but low tubal-rank cannot guarantee the low slice ranks. What's more, our proposed model is non-convex when $0 < p < 1$, which makes it difficult to give a reliable performance guarantee as done in the convex programs, e.g., [22]. Therefore, we give the following Theorem to show that, under a very mild condition, the exact solutions of Eq. (38) are its critical points.

Definition 7: (Tensor Incoherent Condition) [22] Let the skinny t-SVD of a tensor $\mathcal{Z}$ be $\mathcal{U} * \mathcal{S} * \mathcal{V}^*$. $\mathcal{Z}$ is said to satisfy the standard tensor incoherent condition, if there exists μ such that

$$\begin{aligned} \max_{i=1,\ldots,n_1} \|\mathcal{U}^* * \mathfrak{e}_i\|_F &\leq \sqrt{\frac{\mu r}{n_1 n_3}}, \\ \max_{j=1,\ldots,n_2} \|\mathcal{V}^* * \mathfrak{e}_j\|_F &\leq \sqrt{\frac{\mu r}{n_2 n_3}}, \end{aligned} \tag{39}$$

where $\mathfrak{e}_i$ is the $n_1 \times 1 \times n_3$ column basis with $\mathfrak{e}_{i11} = 1$ and $\mathfrak{e}_j$ is the $n_2 \times 1 \times n_3$ column basis with $\mathfrak{e}_{j11} = 1$. r is the tubal-rank of $\mathcal{Z}$, i.e., $\mathrm{rank}_t(\mathcal{Z}) = r$.

Theorem 5: Consider the problem in Eq. (38) with $p_i \geq 1\,(i = 1, \ldots, I)$ and $1/p = \sum_{i=1}^{I} 1/p_i$. Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $n_1 \geq n_2$, $\Omega \sim Ber(\rho)$ and the skinny t-SVD of $\mathcal{T}$ be $\mathcal{U} * \mathcal{S} * \mathcal{V}^*$, and $\mathrm{rank}_t(\mathcal{T}) = r$.

If $\mathcal{T}$ satisfies the Tensor Incoherent Condition with parameter μ, and $\rho \geq \mathcal{O}\left(\mu r \log(n_1 n_3)/(n_1 n_3)\right)$, then with a high possibility, the exact solution of Eq. (38), denoted by $(\widehat{\mathcal{X}}_1, \ldots, \widehat{\mathcal{X}}_I)$:

$$\begin{aligned} \widehat{\mathcal{X}}_1^* &= \mathcal{U} * \mathcal{S}^{p/p_1} * \mathcal{Q}_1^*, \\ \widehat{\mathcal{X}}_i^* &= \mathcal{Q}_{i-1} * \mathcal{S}^{p/p_i} * \mathcal{Q}_i^*,\ i = 2, \ldots, I-1, \\ \widehat{\mathcal{X}}_I &= \mathcal{Q}_{I-1} * \mathcal{S}^{p/p_I} * \mathcal{V}^*, \end{aligned} \tag{40}$$

where $\mathcal{Q}_i \in \mathbb{R}^{q_i \times r \times n_3}$, $\mathcal{Q}_i^* * \mathcal{Q}_i = \mathcal{I}$ for all i and $q_i \geq r$, is a critical point of the problem in Eq. (38).

Theorem 5 gives a new perspective on our non-convex tensor completion problem (38). When a certain optimization procedure converges to a stationary point, it may be close to the exact solutions.

VI. Experiments

In this section, we conduct numerical experiments to evaluate our proposed model. We apply the t-Schatten-p norm (tSp) to solve the TRPCA and the TC problems. The results on both synthetic and real-world datasets demonstrate the superiority of our method. The numbers reported in all the experiments are averaged from 20 random trials.

In [4], Zhang *et al.* set $\lambda = \frac{n_3}{\sqrt{\max\{n_1, n_2\}}}$. In [3], Lu *et al.* set $\lambda = \frac{1}{\sqrt{\max\{n_1, n_2\} \times n_3}}$ and give some good properties. Because TNN is the main method we need to compare with, we extend its choice of λ to our multi-factors so that it can be regarded as a special case of our method. In the following experiments, we usually set the parameter $\lambda = \sqrt{\frac{I}{\max\{n_1, n_2\} \times n_3}}$ in Eq. (41) and Eq. (42), where I denotes the number of factors and data $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. But sometimes we need to readjust λ around the default value for a better experimental result.

A. Tensor Robust Principal Component Analysis

1) Model and Experimental Settings: For the TRPCA problem, Ψ in Eqs. (19) or (20) is an indentity operator. Then the

TRPCA model based on our t-Schatten-p norm is as follows:

$$\min_{\{\mathcal{X}_i\},\mathcal{E}} \quad \sum_{i=1}^{I} \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i} + \lambda\|\mathcal{E}\|_1,$$
$$\text{s.t.} \quad \mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I + \mathcal{E} = \mathcal{T}. \tag{41}$$

Data: To show the advantages of the proposed method, we experiment with both synthetic and real data. The real datasets cover one computer vision task: sequential face images denoising.[5]

Baseline: In this part, we compare our method with TNN based [3] and SNN based [36] methods. These two methods are widely used in various applications.

Evaluation metrics: Assume that the clean tensor is $\mathcal{X}_0 \in \mathbb{R}^{n_1\times n_2\times n_3}$, we represent the recovered tensor (the output of the algorithms) as $\mathcal{X}$.

– *Relative Square Error* (RSE): The reconstruction error is computed as:

$$RSE = \frac{\|\mathcal{X}_0 - \mathcal{X}\|_F}{\|\mathcal{X}_0\|_F}.$$

– *Peak Signal-to-Noise Ratio* (PSNR):

$$PSNR = 10\log_{10}\left(\frac{n_1 n_2 n_3\|\mathcal{X}_0\|_\infty^2}{\|\mathcal{X} - \mathcal{X}_0\|_F^2}\right).$$

2) Synthetic Experiments: We only compare our methods with the TNN based method [3] on the synthetic dataset, because both of our methods are used to solve the low tubal-rank minimization problems, yet the SNN method unfolds the tensor into matrices along each dimension and minimizes the Tucker-n-rank [13].

Here we firstly generate a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ ($n_1 = n_2 = n_3 = 50$) with each entry coming from the normal distribution $\mathcal{N}(0,1)$, then we obtain a low tubal-rank tensor by truncating the singular values vectors in the frequency domain. The tubal-rank is set to 20. For generating the noise/outliers tensor $\mathcal{E}$, we create an index set Ω by using a Bernoulli model to randomly sample a subset from $\{1,\ldots,n_1\} \times \{1,\ldots,n_2\} \times \{1,\ldots,n_3\}$. The noise/outliers fraction is 0.1 here with each entry of the tensor obeying the distribution $\mathcal{N}(0,3)$ if its index is contained in the index set Ω.

Fig. 2 shows the RSEs of the competing methods with respect to the iteration steps. We compare our methods with different selections of $\mathbf{p}$ with TNN, where $\mathbf{p}$ denotes the vector consisting of all p_i's. Since the optimization problem of TNN is convex and easy to solve, TNN converges faster than our methods. However, our methods can exactly recover the underlying low tubal-rank tensor. This is because we utilize a tighter rank approximation of each front slice of the Fourier transformed tensor. Although $p < 1$ makes the optimization non-smooth and non-convex, with the help of Theorem 1, we can still solve the optimization problem efficiently and exactly. For triple-fraction $\mathbf{p} = [2,2,2]$, we found that it has a slower convergence rate than the double-fraction case ($\mathbf{p} = [1,1],[1,2]$). Although its subproblem is smooth and convex, the triple-factor reformulation makes the objective surface more complex, which brings more difficulties to optimization.

[5]Obtained from http://www.cs.nyu.edu/~roweis/data.html

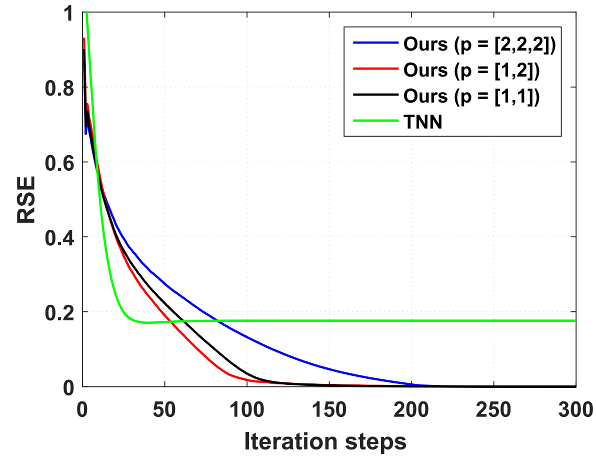


Fig. 2. The convergence of the competing methods on the TRPCA problem.

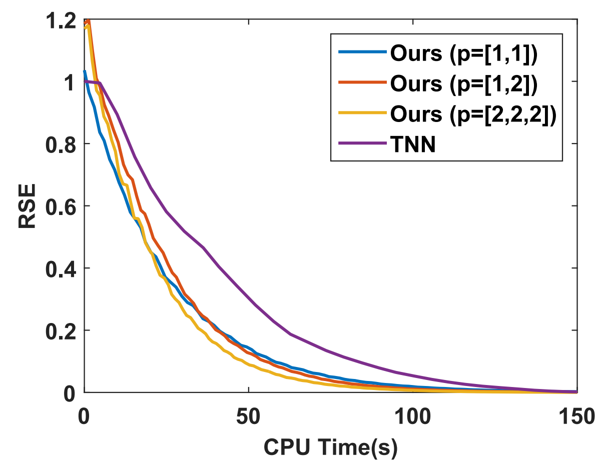


Fig. 3. RSEs of the competing methods vs. CPU time.

Fig. 3 shows the RSEs of the competing methods with respect to CPU times. We generate a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ ($n_1 = n_2 = 800$, $n_3 = 10$) with $\text{rank}_t(\mathcal{X}) = 20$ and noise fraction setting to 10%. The results show that when the tubal-rank is much lower than $\min\{n_1,n_2\}$, due to the smaller computational complexities, our methods are more efficient than TNN.

Combining the results of Fig. 2 and Fig. 3, in order to obtain a faster convergence rate, we choose $\mathbf{p} = [1,2]$ for the following experiments.

For further comparison, we firstly generate a tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ ($n_1 = n_2 = n_3 = 100$) and then change the tubal-rank from 5 to 43 and vary the noise/outliers fraction $|\Omega|/(n_1 n_2 n_3)$ from 0.05 to 0.45 with a step size equaling 0.02. Fig. 4 compares the proposed method with the convex TNN method. Not surprisingly, in terms of the number of successfully restored matrices, our methods outperforms TNN by 40% around. And from the results in Fig. 4, our method is much more robust to the noise/outliers in the relatively high rank case and also performs well when the noise rate is also high, which coincides with the similar phenomenon in the matrix case [25].

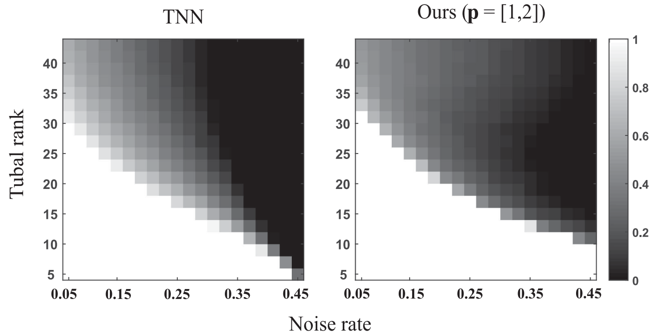
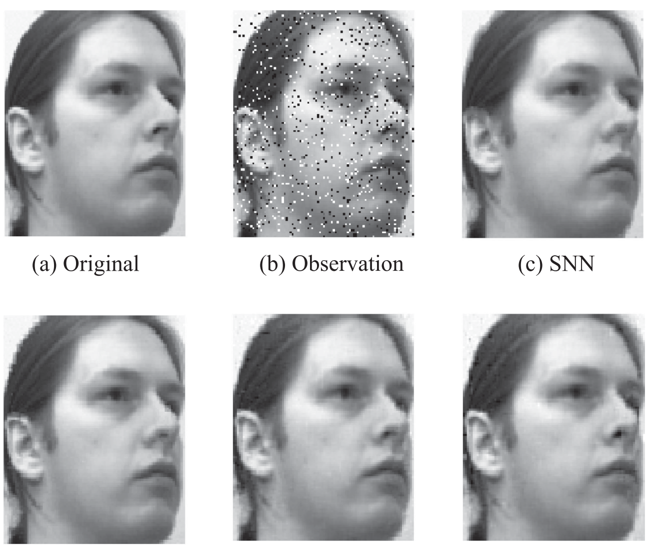


Fig. 4. Comparing our method with convex optimization with TNN. The numbers plotted on the above figures are the success rates within 30 random trials. The white and black areas mean succeed and fail, respectively. Here, the success is in a sense that PSNR $\geq$ 40 dB.



(a) Original (b) Observation (c) SNN

(d) TNN (e) Ours1 (f) Ours2

Fig. 5. Examples of face image denoising. (a) the original image. (b) the observed image. (c)–(f) the denoising results of SNN, TNN, Ours1 ($\mathbf{p} = [1, 2]$), and Ours2 ($\mathbf{p} = [2, 2, 2]$), respectively.

Actually, the conditions of the performance guarantee for the non-convex model are weaker than the convex one, which bring the benefits to the t-Schatten-p norm for the TRPCA problem.

3) Image Denoising: We compare the methods on the face image denoising problem. This face dataset consist of 575 grayscale face images with a size of 112×92. All the entries of the tensor are scaled to $[0, 1]$ and the noise tensor is generated the same as that in the synthetic case with entry from the distribution $\mathcal{N}(0, 3)$. Since the images for one individual are cropped from different views, it is actually a high rank matrix if we vectorize the images and concatenate them as a matrix. Fortunately, as shown in [23], the frontal slice of the Fourier transformed face tensor is low-rank, namely, we can denoise the face images by pursing the low tubal-rank structure.

Fig. 5 gives some denoising results of the competing methods with the noise rate equaling 0.1. Our methods can deal with the details (areas near the nose and hair) better than TNN and SNN. For our models and TNN, we all set one penalty coefficient for

TABLE I
COMPARISON OF PSNR RESULTS ON FACE IMAGES WITH DIFFERENT NOISE RATES

Noise Rate	0.1	0.2	0.3	0.4
SNN	28.17	25.48	22.35	21.27
TNN	28.94	26.58	23.92	21.99
Ours ($\mathbf{p} = [1, 1]$)	29.89	**27.55**	25.94	**23.85**
Ours ($\mathbf{p} = [1, 2]$)	**30.41**	26.92	**26.08**	23.10
Ours ($\mathbf{p} = [2, 2, 2]$)	29.92	27.03	25.97	21.96

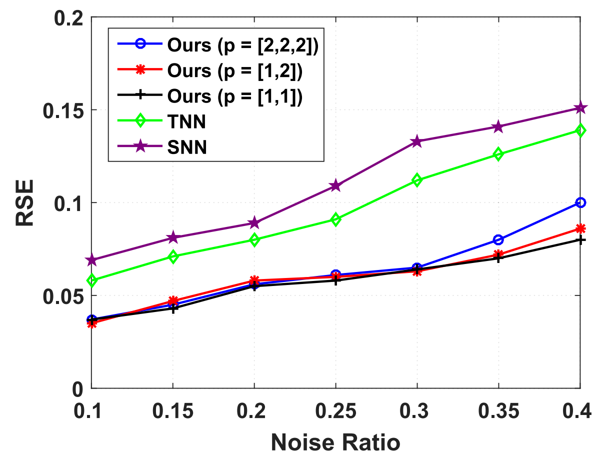


Fig. 6. The RSEs of the competing methods with different noise ratios.

the whole multi-rank vectors. Our t-Schatten-p norm is much tighter than the tensor nuclear norm, therefore our proposed models have lower probability of over penalizing the tensor rank, which can preserve the details of the images. Some numerical results are reported in Table I and Fig. 6.

Table I is the collection of the competing methods' PSNRs. The noise scale is set to 3, i.e., the non-zero entry of the noise tensor is generated from the distribution $\mathcal{N}(0, 3)$. The results show that with the increase of noise rate, the advantages of our methods are more and more obvious. Fig. 6 shows the RSE between the original images and the recovered images, some of which corresponding to the various cases in Table I. Comparing with TNN and SNN, the recovered tensors obtained by our methods are closer to the original data. These results show that our method has a stronger ability to identify the outliers than other two methods, which is very important for some scenarios, such as medical image processing and outlier detection.

B. Tensor Completion

1) Model and Experimental Settings: For the TC problem, Ψ in Eqs. (19) or (20) is an orthogonal projection operator $\mathcal{P}_\Omega$. Then the TC model based on our tSp is given by:

$$\min_{\{\mathcal{X}_i\},\mathcal{E}} \quad \sum_{i=1}^{I} \frac{1}{p_i} \|\mathcal{X}_i\|_{S_{p_i}}^{p_i} + \lambda \|\mathcal{E}\|_F^2,$$
$$\text{s.t.} \quad \mathcal{P}_\Omega(\mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I + \mathcal{E}) = \mathcal{P}_\Omega(\mathcal{T}). \tag{42}$$

Data: We evaluate our method and other state-of-the-art methods on two inpainting tasks: 1) color image inpainting [37]

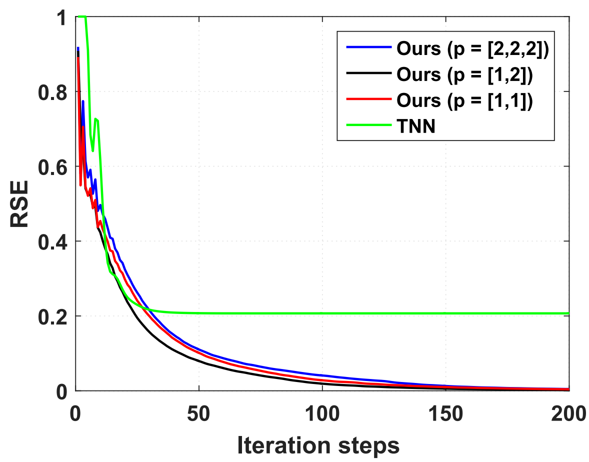


Fig. 7. The convergence of the competing methods on the TC problem.

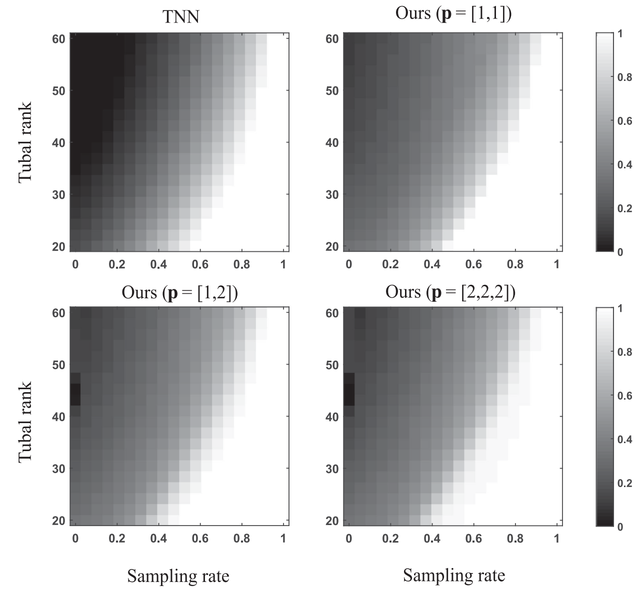


Fig. 8. Comparing our method with convex optimization TNN. The numbers plotted on the above figures are the success rates within 30 random trials. The white and black areas mean succeed and fail, respectively. Here, the success is in a sense that PSNR $\geq$ 40 dB.

(Berkeley Segmentation database); 2) grayscale video inpainting (YUV Video Sequences).

Baseline: In this part we compare our proposed method with other state-of-the-arts, including TMac-inc [38], SiLRTC [13], TCTF [23], and TNN [4] on inpainting applications. The codes are provided by their corresponding authors. Note that our proposed tSp together with TNN and TCTF are all based on t-product, while TMac-inc and SiLRTC are based on Tucker product. They all have their own theoretical guarantees, thus we compare these methods together, but the first three are compared emphatically.

Evaluation metrics: We use the same metrics as the TRPCA case, i.e., PSNR and RSE.

2) Synthetic Experiments: We generate a low-rank tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ $(n_1 = n_2 = n_3 = 100)$ and an index set Ω by the following steps. First, we produce two tensors $\mathcal{A} \in \mathbb{R}^{n_1\times r\times n_3}$ and $\mathcal{B} \in \mathbb{R}^{r\times n_2\times n_3}$. Then let $\mathcal{X} = \mathcal{A} * \mathcal{B}$ to get a tensor with $\text{rank}_t(\mathcal{X}) = r$. After that, we create the index set Ω by using a Bernoulli model to randomly sample a subset from $\{1,\ldots,n_1\} \times \{1,\ldots,n_2\} \times \{1,\ldots,n_3\}$. The sampling rate is $|\Omega|/(n_1n_2n_3)$.

Fig. 7 presents the RSE of TNN and the proposed methods with respect to the iteration steps. Here $\text{rank}_t(\mathcal{X})$ equals to 40 and the sampling rate equals to 0.4. TNN converges much faster than our methods due to its convexity. Another reason is that we adopt the proximal gradient to update the variables while TNN can achieve closed-form solutions for its subproblems. Nevertheless, all our methods with various selections of $\mathbf{p}$ can exactly recover the ground-truth low tubal-rank tensor. Same as TRPCA, this is because the condition of the sampling rate for exact recovery is weaker than that of the convex nuclear norm, namely, the proposed t-Schatten-p norm still works well when the sample size is low.

The exhaustive comparison between TNN and our methods is shown in the Fig. 8. Our method outperforms TNN by 20% around when $\mathbf{p} = [2, 2, 2]$. These results verify the effectiveness of our t-Schatten-p norm. We can see that the performance for $\mathbf{p} = [1, 2]$ and $\mathbf{p} = [2, 2, 2]$ are similar. This is not strange because they both correspond to the t-Schatten-$2/3$ norm.

3) Image Inpainting: We use the Berkeley Segmentation dataset[6] [37] to evaluate our method for the image inpainting task. This dataset has totally 200 RGB images, each with size $321 \times 481 \times 3$. As pointed out by [23], most natural images have the low tubal-rank structure, therefore we can inpaint these natural images by the low-rank tensor completion method.

Fig. 9 gives the visualizations of the image inpainting on some images with 0.4 sampling rate. Our methods have the highest PSNR among all the competing completion methods. We choose $\mathbf{p} = [1, 2]$ for Ours1 and $\mathbf{p} = [2, 2, 2]$ for Ours2. According to the results, our methods can deal with the details of images better than TNN. The reason is that our t-Schatten-p norm is much tighter than the tensor nuclear norm in approximating the tensor tubal-rank. Thus the t-Schatten-p norm has a smaller possibility of over penalizing the singular values of each frontal slice of the Fourier transformed tensor. Therefore, the t-Schatten-p norm based method is more suitable for inpainting images with complex details.

Table II shows the average PSNR results on 40 natural images, randomly chosen from the dataset, with different sampling rates. It is obvious that our methods still work well when the sampling rate is very low (rate $= 0.2$). Our methods outperform the others much more when rates are less than 0.6. Note that although all the competing methods have similar results when the sampling rate is high, our methods are more memory saving than the others due to the factorization strategy.

4) Video Inpainting: We evaluate our methods on the widely used YUV Video Sequences.[7] Each sequence contains at least 150 frames. In the experiments, we test our methods and other

[6]Obtained from https://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/

[7]Obtained from http://trace.eas.asu.edu/yuv/

(a) Original (b) Observation (c) SiLRTC (d) TCTF (e) TMAC (f) TNN (g) Ours1 (h) Ours2

Fig. 9. Examples of image inpainting. (a) the original image. (b) the observed image. (c)–(h) the inpainting results of SiLRTC, TCTF, TMac-inc, TNN, Ours1 ($\mathbf{p} = [1, 1]$), and Ours2 ($\mathbf{p} = [2, 2, 2]$), respectively.

TABLE II
COMPARISON OF PSNR RESULTS ON NATURAL IMAGES WITH DIFFERENT SAMPLING RATES

Sampling Rate	0.2	0.4	0.6	0.8
TMac-inc	18.72	24.18	28.55	38.74
SiLRTC	20.36	25.37	29.20	39.30
TCTF	21.11	27.10	29.33	39.62
TNN	23.10	27.98	33.29	41.39
Ours ($\mathbf{p} = [1, 1]$)	**25.29**	**29.85**	34.42	**42.08**
Ours ($\mathbf{p} = [1, 2]$)	24.41	28.72	33.68	41.31
Ours ($\mathbf{p} = [2, 2, 2]$)	24.11	28.77	**34.97**	41.57

methods on two videos. The frame sizes of the two videos are both 144×176 pixels. The work in [23] reveals that the tensor of grayscale video has much redundant information because of its similar contents within and between frames, and thus its low tubal-rank structure is notable. We can complete the missing entries of the tensor by tensor low-rank minimization.

Due to the computational limitation, we only use the first 30 frames of the two sequences. As shown in Fig. 11, we display the 10-th frame of the two testing videos, respectively. From the recovery results, our methods perform better in filling the missing values of the two video sequences. It can deal with the details better.

Table III shows the PSNR metric of the competing methods. Our methods achieve the best inpainting recovery, consistent with the observations in Fig. 11. Low tubal-rank methods (TNN and Ours) are also better than the others, which demonstrate that the low tubal-rank structure does benefit the tensor completion task on video sequence. Comparing to TNN, our methods can maintain the details of the video sequence better. This is because, one hand, the t-Schatten-p norm is much tighter than the nuclear norm for approximating the tensor multi-rank. On the other hand, the surrogate of the t-Schatten-p norm introduces the convexity and smoothness to the subproblems of the optimization, which will reduce the disadvantages of setting $p < 1$ for our norm. Hence, by Theorem 5 we can still achieve a good stationary point.

TABLE III
COMPARISON OF PSNR RESULTS ON VIDEO INPAINTINGS WITH DIFFERENT SAMPLING RATES

Hall Monitor				
Sampling rate	0.1	0.2	0.4	0.6
TMac-inc	16.06	22.29	27.55	38.74
SiLRTC	19.23	21.64	28.47	39.30
TCTF	19.32	23.08	28.50	39.62
TNN	20.00	25.00	29.33	41.39
Ours ($\mathbf{p} = [1, 1]$)	22.08	27.32	31.23	41.31
Ours ($\mathbf{p} = [1, 2]$)	**23.71**	**28.14**	**32.14**	**42.08**
Ours ($\mathbf{p} = [2, 2, 2]$)	23.02	27.92	31.75	41.57
Akiyo				
Sampling rate	0.1	0.2	0.4	0.6
TMac-inc	15.98	21.06	26.41	28.99
SiLRTC	18.52	23.93	27.12	28.60
TCTF	20.11	22.07	28.09	29.37
TNN	19.97	24.20	29.60	33.43
Ours ($\mathbf{p} = [1, 1]$)	21.10	26.39	30.42	34.02
Ours ($\mathbf{p} = [1, 2]$)	**23.79**	**26.86**	**32.09**	**35.68**
Ours ($\mathbf{p} = [2, 2, 2]$)	23.16	26.02	31.86	35.03

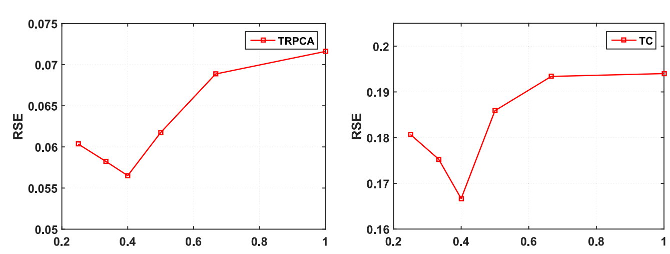


Fig. 10. RSEs of the TC and the TRPCA problems with respect to various selections of p for the t-Schatten-p norm.

C. Discussion on the Choice of p

In this section, we study the relation between the performance and the value of p for our t-Schatten-p norm. We generate a low tubal-rank tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ $(n_1 = n_2 = n_3 = 50)$ with

(a) Original (b) Observation (c) SiLRTC (d) TCTF (e) TMAC (f) TNN (g) Ours1 (h) Ours2

Fig. 11. Examples of video inpainting. (a) the original image. (b) the observed image. (c)–(h) the inpainting results of SiLRTC, TCTF, TMac-inc, TNN, Ours1 ($\mathbf{p} = [1, 1]$), and Ours2 ($\mathbf{p} = [2, 2, 2]$), respectively.

$\text{rank}_t(\mathcal{X}) = 20$ and the noise rate and sampling rate equaling 0.4. For each value of $p \in \{1/4, 1/3, 2/5, 1/2, 2/3, 1\}$, each experiment is repeated 20 times and the average RSEs are reported in Fig. 10, from which we can see that the RSEs increase when the value of p rises in the range $[2/5, 1]$. This result clearly justifies the validity of our proposed t-Schatten-p norm for solving the LRTR problems when $p < 1$. Actually, a smaller p represents a tighter approximation to the tensor tubal-rank. Note that when $p = 0$, the t-Schatten-p norm reduces to the tensor rank function. However, a smaller p makes the objective function more non-convex and non-smooth and thus more difficult to optimize. And accoarding to Theorem 1, a smaller p (in the range $(0, 2/5)$) may require more tensor factors, and tensor multiplication makes the problem (20) non-convex, which may lead to a bad solution. Besides, the equivalence condition of Theorem 1 for each factor $\mathcal{X}_i$ is not a single point. Instead, each $\mathcal{X}_i$ belongs to a large subset by multiplying unitary tensors, which also increases the difficulty of optimization.

VII. Conclusions

We propose a new definition of tensor Schatten-p norm named as the t-Schatten-p norm. When $p < 1$, our t-Schatten-p norm can better approximate the ℓ_1 norm of tensor multi-rank than TNN. Therefore, we use this norm to solve the LRTR problem as a tighter regularizer. We further provide the surrogate theorem for our proposed t-Schatten-p norm and give an efficient algorithm to solve the LRTR problem. We also provide some theoretical analysis on exact recovery and the corresponding error bound for the noise case. The experimental results on TRPCA and TC show that our methods perform better than the mainstream methods when the clean data have a large tubal-rank or a high noise/corruption ratio. Finally, We also discuss the choice of p, and recommend a range for selecting p for the LRTR problem.

Appendix

A. Supplementary Definition

The followings are the definitions of tensor transpose, identity tensor, and orthogonal tensor, respectively.

Definition 8: (Tensor transpose) [1] The conjugate transpose of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the $\mathcal{T}^* \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ obtained by conjugate transposing each frontal slice of $\overline{\mathcal{T}}$, i.e.,

$$\mathcal{T}^* \text{is the transpose of } \mathcal{T} \iff \overline{\mathbf{T}^*}^{(k)} = \left(\overline{\mathbf{T}}^{(k)}\right)^H. \tag{43}$$

$\mathcal{T}^*$ can also be obtained by conjugate transposing each of $\mathcal{T}$'s frontal slice and then reversing the order of transposed frontal slices 2 through n_3.

Definition 9: (Identity tensor) [1] Let $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$, then $\mathcal{I}$ is an identity tensor if its first frontal slice $\mathbf{I}^{(1)}$ is the $n \times n$ identity matrix and all other frontal slices $\mathbf{I}^{(i)}, i = 2 \ldots, n_3$, are zero matrices.

Definition 10: (Orthogonal tensor) [1] Let $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$, then $\mathcal{Q}$ is orthogonal if it satisfies

$$\mathcal{Q}^* * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}. \tag{44}$$

B. Proof of Theorem 2

Proof: For convenience, we let the variables without and with superscript $+$ represent the variable in the k and $k+1$ iteration variable and $\|\cdot\|$ denote any well-defined matrix/tensor norm. Assume we already have $\|\mathcal{Y}^+ - \mathcal{Y}\| \to 0$ and $\|\mathcal{Z}^+ - \mathcal{Z}\| \to 0$, then $\|\nabla_{\mathcal{Z}}\mathcal{L}\| = \frac{1}{\rho_2}\|\mathcal{Z}^+ - \mathcal{Z}\| = \|\mathcal{X}_1 * \mathcal{X}_2 * \cdots * \mathcal{X}_I - \mathcal{G}\| \to 0$ and $\|\nabla_{\mathcal{Y}}\mathcal{L}\| = \frac{1}{\rho_1}\|\mathcal{Y}^+ - \mathcal{Y}\| = \|\Psi(\mathcal{G}) + \mathcal{E} - \mathcal{T}\| \to 0$, hence the equality constraint is satisfied at the limit points. Since our optimization problem is a multi-linear optimization, by Corollary 6.21 in [39], if we have $\|\mathcal{X}_i^+ - \mathcal{X}_i\|$, $\|\mathcal{G}^+ - \mathcal{G}\|$, and $\|\mathcal{E}^+ - \mathcal{E}\|$ all approaching 0, then there exists $\upsilon \in \partial\mathcal{L}$ with $\upsilon \to 0$. Hence, any limit point produced by our algorithm is a constrained stationary point. We now prove the convergence of the differences to 0.

Lemma 2: The change in the augmented Lagrangian when the primal variable $\mathcal{X}_i$ is updated to $\mathcal{X}_i^+$ is given by

$$\begin{aligned}
&\mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) \\
&= \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i} - \frac{1}{p_i}\|\mathcal{X}_i^+\|_{S_{p_i}}^{p_i} - \langle \upsilon, \mathcal{X}_i - \mathcal{X}_i^+ \rangle \\
&+ \rho_2 L_i \|\mathcal{X}_i - \mathcal{X}_i^+\|_F^2 + \frac{\rho_2}{2}\|C(\mathcal{X}_i) - C(\mathcal{X}_i^+)\|_F^2,
\end{aligned} \tag{45}$$

where $C(\mathcal{X}_i) = \mathcal{Q}_{1:(i-1)} * \mathcal{X}_i * \mathcal{Q}_{(i+1):I} - \mathcal{G}$ is a linear transformation and $\upsilon \in \partial(\frac{1}{p_i}\|\mathcal{X}_i^+\|_{S_{p_i}}^{p_i})$.

Proof: Expanding $\mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z})$, the change is

$$\frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i} - \frac{1}{p_i}\|\mathcal{X}_i^+\|_{S_{p_i}}^{p_i} + \langle \mathcal{Z}, C(\mathcal{X}_i) - C(\mathcal{X}_i^+)\rangle$$
$$+ \frac{\rho_2}{2}(\|C(\mathcal{X}_i) - \mathcal{G}\|_F^2 - \|C(\mathcal{X}_i^+) - \mathcal{G}\|_F^2)$$
$$= \frac{1}{p_i}\|\mathcal{X}_i\|_{S_{p_i}}^{p_i} - \frac{1}{p_i}\|\mathcal{X}_i^+\|_{S_{p_i}}^{p_i} + \frac{\rho_2}{2}\|C(\mathcal{X}_i) - C(\mathcal{X}_i^+)\|_F^2$$
$$+ \langle \mathcal{Z}, C(\mathcal{X}_i) - C(\mathcal{X}_i^+)\rangle + \rho_2\langle C(\mathcal{X}_i) - C(\mathcal{X}_i^+), C(\mathcal{X}_i^+) - \mathcal{G}\rangle.$$

We observe that

$$\langle \mathcal{Z}, C(\mathcal{X}_i) - C(\mathcal{X}_i^+)\rangle + \rho_2\langle C(\mathcal{X}_i) - C(\mathcal{X}_i^+), C(\mathcal{X}_i^+) - \mathcal{G}\rangle$$
$$= \langle \mathcal{Z} + \rho_2(C(\mathcal{X}_i^+) - \mathcal{G}), C(\mathcal{X}_i) - C(\mathcal{X}_i^+)\rangle$$
$$= \langle C^T(\mathcal{Z} + \rho_2(C(\mathcal{X}_i^+) - \mathcal{G})), \mathcal{X}_i - \mathcal{X}_i^+\rangle.$$

Note that $C^T(\mathcal{Z} + \rho_2(C(\mathcal{X}_i^+) - \mathcal{G})) = \rho_2 \nabla f_i^+(\mathcal{X}_i^+)$. On the other hand, by the optimality of Eq. (23), we have $\rho_2 \nabla f_i^+ (\mathcal{X}_i^+) + \rho_2 L_i(\mathcal{X}_i^+ - \mathcal{X}_i) \in -\partial(\frac{1}{p_i}\|\mathcal{X}_i^+\|_{S_{p_i}}^{p_i})$. Hence, we can obtain Eq. (45) directly.

Due to Lemma 2 and the convexity of the function $\|\cdot\|_{S_{p_i}}^{p_i}$ when $p_i \geq 1$, we have

$$\mathcal{L}(\mathcal{X}_i^+, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) + c_1\|(\mathcal{X}_i^+ - \mathcal{X}_i)\|_F^2 \leq \mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}), \tag{46}$$

where $c_1 = \rho_2 L_i > 0$. Since we update $\mathcal{G}$ by the closed-form solution of Eq. (27), together with the strong convexity of the objective, we have

$$\mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) + \frac{\mu_1}{2}\|\mathcal{G} - \mathcal{G}^+\|_F^2 \leq \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}), \tag{47}$$

where $\mu_1 > 0$. By the same derivation of Lemma 2,

$$\mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}, \mathcal{Z})$$
$$= g(\mathcal{E}) - g(\mathcal{E}^+) - \langle \upsilon, \mathcal{E}_i - \mathcal{E}_i^+\rangle + \frac{\rho_1}{2}\|C(\mathcal{E}_i) - C(\mathcal{E}_i^+)\|_F^2,$$

where $\upsilon = \nabla(g(\mathcal{E}^+))$ and $C(\mathcal{E}) = \Psi(\mathcal{G}) + \mathcal{E} - \mathcal{T}$. Note that $g(\cdot)$ in Theorem 2 is convex in our case, hence we get

$$\mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}, \mathcal{Z}) + \frac{\rho_1}{2}\|\mathcal{E} - \mathcal{E}^+\|_F^2 \leq \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}, \mathcal{Y}, \mathcal{Z}). \tag{48}$$

According to the Eq. (33), it is easy to verify that

$$\mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}^+, \mathcal{Z}^+) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}, \mathcal{Z})$$
$$= \frac{1}{\rho_1}\|\mathcal{Y}^+ - \mathcal{Y}\|_F^2 + \frac{1}{\rho_2}\|\mathcal{Z}^+ - \mathcal{Z}\|_F^2. \tag{49}$$

By the optimality of $\mathcal{E}^+$, we have

$$\mathcal{Y}^+ = \rho_1(\Psi(\mathcal{G}^+) + \mathcal{E}^+ - \mathcal{T}) + \mathcal{Y} = -\lambda\nabla(g(\mathcal{E}^+)). \tag{50}$$

The objective of (27) is differentiable and by the optimality of $\mathcal{G}^+$, we similarly have

$$\Psi(\mathcal{Y}^+) = \mathcal{Z}^+. \tag{51}$$

Therefore, we conclude

$$\mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}^+, \mathcal{Z}^+) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}, \mathcal{Z})$$
$$\overset{(a)}{\leq} \frac{2}{\rho_1}\|\mathcal{Y}^+ - \mathcal{Y}\|_F^2 \overset{(b)}{\leq} \frac{2\lambda^2 c_2}{\rho_1}\|\mathcal{E}^+ - \mathcal{E}\|_F^2, \tag{52}$$

where c_2 is some constant, we assume $\rho_2 \geq \rho_1$ in (a) and use the gradient Lipschitz property of $g(\cdot)$ for (b). Combing the Eq. (46) and (52), we find that the augmented Lagrangian function is monotonically decreasing. The function $\mathcal{L}$ is upper bounded. It is easy to verify the Lagrangian function $\mathcal{L}$ is also lower bounded and coercive. Thus, all the variables produced by our algorithm are bounded. We also obtain

$$\mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}^+, \mathcal{Z}^+)$$
$$\geq c_1\|(\mathcal{X}_i^+ - \mathcal{X}_i)\|_F^2 + \frac{\mu_1}{2}\|\mathcal{G} - \mathcal{G}^+\|_F^2$$
$$+ \left(\frac{\rho_1}{2} - \frac{2\lambda c_2}{\rho_1}\right)\|\mathcal{E} - \mathcal{E}^+\|_F^2. \tag{53}$$

With the proper choice of ρ_1 and ρ_2 such that $(\frac{\rho_1}{2} - \frac{2\lambda c_2}{\rho_1}) > 0$, we can derive

$$\sum_{k=1}^{\infty}(c_1\|(\mathcal{X}_i^+ - \mathcal{X}_i)\|_F^2 + c_3\|\mathcal{G} - \mathcal{G}^+\|_F^2 + c_4\|\mathcal{E} - \mathcal{E}^+\|_F^2$$
$$\leq \sum_{k=1}^{\infty}(\mathcal{L}(\mathcal{X}_i, \mathcal{G}, \mathcal{E}, \mathcal{Y}, \mathcal{Z}) - \mathcal{L}(\mathcal{X}_i^+, \mathcal{G}^+, \mathcal{E}^+, \mathcal{Y}^+, \mathcal{Z}^+)) < \infty, \tag{54}$$

where $c_3 = \frac{\mu_1}{2} > 0$ and $c_4 = \frac{\rho_1}{2} - \frac{2\lambda c_2}{\rho_1} > 0$. We conclude that $\|\mathcal{X}_i^+ - \mathcal{X}_i\| \to 0$, $\|\mathcal{G}^+ - \mathcal{G}\| \to 0$, and $\|\mathcal{E}^+ - \mathcal{E}\| \to 0$. Based on Eqs. (50) and (51) and $\|\mathcal{E}^+ - \mathcal{E}\| \to 0$, we can have $\|\mathcal{Y}^+ - \mathcal{Y}\| \to 0$ and $\|\mathcal{Z}^+ - \mathcal{Z}\| \to 0$. The proof is finished. ■

Acknowledgment

The authors would like to thank J. Wu for helping us improve the manuscript.

References

[1] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Its Appl.*, vol. 435, no. 3, pp. 641–658, 2011.

[2] 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.*, vol. 34, no. 1, pp. 148–172, 2013.

[3] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2016, pp. 5249–5257.

[4] 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 *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2014, pp. 3842–3849.

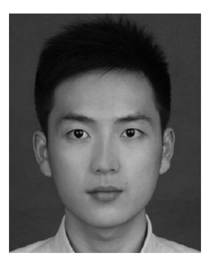
[5] Y. Fu, J. Gao, D. Tien, Z. Lin, and X. Hong, "Tensor LRR and sparse coding-based subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 2120–2133, Oct. 2016.

[6] N. Boumal and P.-a. Absil, "RTRMC: A Riemannian trust-region method for low-rank matrix completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 406–414.

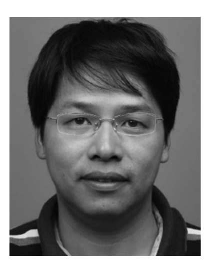
[7] B. Romera-Paredes, H. Aung, N. Bianchi-Berthouze, and M. Pontil, "Multilinear multitask learning," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1444–1452.

[8] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[9] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Studies Appl. Math.*, vol. 6, no. 1–4, pp. 164–189, 1927.
[10] H. A. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.
[11] J. B. Kruskal, *Rank, Decomposition, and Uniqueness For 3-Way and n-Way Arrays*. Amsterdam, The Netherlands: North Holland, 1989.
[12] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
[13] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
[14] S. Friedland and L.-H. Lim, "Nuclear norm of higher-order tensors," *Math. Comput.*, vol. 87, no. 311, pp. 1255–1281, 2018.
[15] M. Yuan and C.-H. Zhang, "On tensor completion via nuclear norm minimization," *Found. Comput. Math.*, vol. 16, no. 4, pp. 1031–1068, 2016.
[16] J. Håstad, "Tensor rank is NP-complete," *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.
[17] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *J. ACM*, vol. 60, no. 6, 2013, Art. no. 45.
[18] H. Kasai and B. Mishra, "Low-rank tensor completion: A Riemannian manifold preconditioning approach," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1012–1021.
[19] C. Li, L. Guo, Y. Tao, J. Wang, L. Qi, and Z. Dou, "Yet another Schatten norm for tensor recovery," in *Proc. Int. Conf. Neural Inf. Process.*, 2016, pp. 51–60.
[20] B. Romera-Paredes and M. Pontil, "A new convex relaxation for tensor completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2967–2975.
[21] R. Tomioka and T. Suzuki, "Convex tensor decomposition via structured Schatten norm regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1331–1339.
[22] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2017.
[23] P. Zhou, C. Lu, Z. Lin, and C. Zhang, "Tensor factorization for low-rank tensor completion," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1152–1163, Mar. 2018.
[24] X.-Y. Liu, S. Aeron, V. Aggarwal, and X. Wang, "Low-tubal-rank tensor completion using alternating minimization," *Proc. SPIE*, vol. 9848, 2016, Art. no. 984809.
[25] C. Xu, Z. Lin, and H. Zha, "A unified convex surrogate for the Schatten-p norm," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 926–932.
[26] F. Shang, J. Cheng, Y. Liu, Z.-Q. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust PCA: Algorithms and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2066–2080, Sep. 2018.
[27] N. Srebro, J. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 1329–1336.
[28] F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient Schatten p-norm minimization," in *Proc. AAAI Conf. Artif. Intell.*, 2012, pp. 655–661.
[29] G. Liu, Q. Liu, and X. Yuan, "A new theory for matrix completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 785–794.
[30] Z. Lin, R. Liu, and H. Li, "Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning," *Mach. Learn.*, vol. 99, no. 2, pp. 287–325, 2015.
[31] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2015.
[32] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proc. IEEE Int. Conf. Comput. Vision*, 2013, pp. 217–224.
[33] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leqslant 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395–407, 2009.
[34] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi, "A simplified approach to recovery conditions for low rank matrices," in *Proc. IEEE Int. Symp. Inf. Theory Proc.*, 2011, pp. 2318–2322.
[35] F. Shang, Y. Liu, and J. Cheng, "Tractable and scalable Schatten quasi-norm approximations for rank minimization," in *Proc. Artif. Intell. Statist.*, 2016, pp. 620–629.
[36] B. Huang, C. Mu, D. Goldfarb, and J. Wright, "Provable models for robust low-rank tensor completion," *Pacific J. Optim.*, vol. 11, no. 2, pp. 339–364, 2015.
[37] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vision*, vol. 2, 2001, pp. 416–423.
[38] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *Inverse Problems Imag.*, vol. 9, no. 2, pp. 601–624, 2017.
[39] W. Gao, D. Goldfarb, and F. E. Curtis, "ADMM for multiaffine constrained optimization," in 2018, arXiv:1802.09592.

Hao Kong received the bachelor's degree in computer science from the University of Science and Technology Beijing in 2016. He is currently working toward the Ph.D. degree in the Key Laboratory of Machine Perception, School of Electronics Engineering and Computer Science, Peking University, Beijing, China. His research interests include machine learning, pattern recognition, and computer vision.



Xingyu Xie (S'17) received the bachelor's degree in automation from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2016. He is currently working toward the master's degree at the NUAA. His research interests include machine learning and computer vision.



Zhouchen Lin (M'00–SM'08–F'18) is a professor with the Key Laboratory of Machine Perception, School of Electronics Engineering and Computer Science, Peking University. His research interests include computer vision, image processing, machine learning, pattern recognition, and numerical optimization. He is an area chair of CVPR 2014/2016/2019, ICCV 2015, NIPS 2015/2018 and AAAI 2019, and a senior program committee member of AAAI 2016/2017/2018 and IJCAI 2016/2018. He is an associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence and the International Journal of Computer Vision. He is a Fellow of IAPR and IEEE.