This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Tensor Recovery With Weighted Tensor Average Rank

Xiaoqin Zhang, *Senior Member, IEEE*, Jingjing Zheng, Li Zhao, Zhengyuan Zhou, and Zhouchen Lin, *Fellow, IEEE*

***Abstract*— In this article, a curious phenomenon in the tensor recovery algorithm is considered: can the same recovered results be obtained when the observation tensors in the algorithm are transposed in different ways? If not, it is reasonable to imagine that some information within the data will be lost for the case of observation tensors under certain transpose operators. To solve this problem, a new tensor rank called weighted tensor average rank (WTAR) is proposed to learn the relationship between different resulting tensors by performing a series of transpose operators on an observation tensor. WTAR is applied to three-order tensor robust principal component analysis (TRPCA) to investigate its effectiveness. Meanwhile, to balance the effectiveness and solvability of the resulting model, a generalized model that involves the convex surrogate and a series of nonconvex surrogates are studied, and the corresponding worst case error bounds of the recovered tensor is given. Besides, a generalized tensor singular value thresholding (GTSVT) method and a generalized optimization algorithm based on GTSVT are proposed to solve the generalized model effectively. The experimental results indicate that the proposed method is effective.**

***Index Terms*— Low-rank recovery, tensor average rank, tensor robust principal component analysis (TRPCA).**

I. Introduction

WITH the rapid advances of data-intensive applications in various engineering and scientific fields, there is a growing explosion of high-dimensional data, including images and videos, which are difficult to store, transmit, and process. Therefore, exploiting low-dimensional structures in such high-dimensional data are increasingly important for understanding such complicated data, and many methods have been proposed [1]–[9].

Among them, principal component analysis (PCA) [7], [8] was first proposed, and it is widely used for dimension reduction and data analysis. In PCA, the Frobenius norm is imposed on the matrix denoting the noise within the data to characterize the magnitude of small noise perturbation. By using the Frobenius norm, the principal components are robust to small noise perturbation, but they are sensitive to gross sparse errors. Thus, PCA fails to work when the data are corrupted by gross sparse errors [3], [4]. To solve this issue, robust PCA (RPCA) [3], [4] is proposed, in which ℓ_0 norm (i.e., the number of nonzero entries in a matrix) is imposed on a sparse matrix denoting the gross sparse errors within the data to characterize the quantity of gross sparse errors.

Currently, most visual data, such as color images and videos, are in the form of tensors. To use the above low-rank matrix recovery methods, these data need to be transformed into 2-D matrices first. However, as [10] points out, the important structures will be lost when a higher-order tensor is transformed into a 2-D matrix. Therefore, many low-rank tensor recovery methods have been proposed in recent years [10]–[18]. A key challenge in low-rank tensor recovery is to define tensor rank appropriately. Traditionally, there are three types of defining methods for tensor rank, including the methods based on CP (CANDECOMP/PARAFAC) decomposition [19], the methods based on Tucker decomposition [10], [11], [19]–[22], and the methods based on some new tensor products [12], [23], [24]. Similar to the definition of matrix rank, the type of defining methods based on CP decomposition defines the rank of a tensor as the minimum number of rank-one decompositions of the given tensor. However, the minimization problem is NP-hard, which restricts the application of this type of method. The second type of defining method is based on the unfolding matrices of the tensor, and they apply the existing matrix recovery theory to the corresponding tensor recovery problem. Thus, this type of method is more popular than the first one. For example, Gandy *et al.* [11] take the sum of the ranks of different unfolding matrices as the rank of the tensor data, and apply the defined tensor rank into tensor completion (TC) problem, the goal of which is to recover an underlying tensor data from its incomplete observations effectively. Since introducing the sum of the ranks of different unfolding matrices will lead to an NP-hard optimization problem, to improve solution efficiency, this sum is approximated by the sum of nuclear norms (SNNs) in [11]. However, SNN is not the convex envelope of the sum of the ranks as stated in [12]. Based on a more balanced matricization, a weighted sum of the ranks of the unfolding matrices is introduced by Liu *et al.* [10]. Similar to [11], the weighted SNNs is adopted in [10] for an efficient solution of the resulting optimization

Manuscript received February 15, 2021; revised November 30, 2021; accepted June 7, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61922064, Grant U2033210, Grant 61731018, and Grant 12026606. *(Corresponding author: Xiaoqin Zhang.)*

Xiaoqin Zhang, Jingjing Zheng, and Li Zhao are with the College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China (e-mail: zhangxiaoqinnan@gmail.com; jjzheng233@gmail.com; lizhao@wzu.edu.cn).

Zhengyuan Zhou is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: zyzhou@stanford.edu).

Zhouchen Lin is with the Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: zlin@pku.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3182541.

Digital Object Identifier 10.1109/TNNLS.2022.3182541

2162-237X © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

model. Furthermore, Zhang *et al.* [13] proposed a general low-rank discovery framework to deal with some unknown transformation and gross sparse errors in the tensor data, in which the weighted sum of Schatten p-norms of the unfolding matrices is used instead of the weighted SNNs. In addition, Zhang *et al.* [13] also provide a proximal gradient-based algorithm with global optimality convergence guarantees to solve the proposed framework for $p = 1$, effectively. Note that, for the weighted sum of ranks-based methods, the weights play an important role, and it is unknown what is the best choice if without any prior. To solve this problem, a new TC method based on the maximum rank of a set of unfolding matrics is proposed to promote the low-rankness of unfolding matrics of the recovered tensor [25]. In addition, since the tensor recovery methods based on the weighted sum of ranks suffer from the high computation cost of the computing of singular value decompositions (SVDs) for the large unfolding matrices, an efficient matrix factorization method for tensor recovery is developed in [26]. Recently, some new tensor ranks based on the tensor product (t-product) have received more and more attention. A tensor tubal rank based on t-product is proposed in [23] and applied to tensor recovery. Besides, Lu *et al.* [12] defined a new tensor rank based on the product (tensor average rank), and proposed a tensor robust principal component analysis (TRPCA) model. The corresponding recovery guarantee for TRPCA is also presented. Considering the effectiveness and wide application of the product-based tensor rank in computer vision, this article focuses on studying the defining method based on t-product.

Although the tensor ranks based on t-product are effective and widely used, there are still a few limitations.

1) The tensor tubal rank is based on the discrete Fourier transformation (DFT) in the 3rd dimension of the tensor. As a result, the tensor tubal ranks of the resulting tensors obtained by performing different transpose operators on the tensor may be different, which may lead to the tensor recovery results relying on the transpose operators. In this article, this issue is referred to as transpose variability of tensor recovery (TVTR). A tensor recovery algorithm has TVTR property if the results of the algorithm are relying on the transpose operators. It can be reasonably imagined that some information within tensor data (the relationship of various views and low-rank prior information from different directions of tensor data) will be lost if only one dimension is considered in a tensor recovery algorithm with TVTR property.
2) Although it is proven that the true value of the models can be exactly recovered under certain conditions for TRPCA based on ℓ_1 norm (i.e., relax ℓ_0 norm and rank function to ℓ_1 norm and nuclear norm, respectively.) These strong conditions often cannot be guaranteed in the real world.

A. Our Contributions

To overcome the aforementioned limitations, this article focuses on the recovery of a low-rank tensor from a three-order data tensor contaminated by both gross sparse errors and small entry-wise dense noise. The contributions of this work are threefold. First, to our best knowledge, TVTR is first discussed in this article. Second, to deal with TVTR, a new tensor rank called weighted tensor average rank (WTAR) is given. Meanwhile, WTAR is applied to the tensor-robust principal component analysis, and a new low-rank tensor recovery model called tensor recovery based on WTAR (TRWTAR) is obtained. In addition, we prove that the worst case error bounds of the recovered tensor are established by TRWTAR (in Theorem 4). Third, inspired by the literature on nonconvex optimization [27]–[32] (see Table I), this article provides a general algorithm that solves both the convex surrogate and a series of nonconvex surrogates of the proposed framework (not limited to the surrogate functions in Table I). The study results contribute to the broad landscape of tensor recovery by delineating an effective measure of tensor rank and providing theoretical and algorithmic advances in robust tensor recovery problems.

TABLE I
SOME SURROGATE FUNCTIONS OF ℓ_0

Name	$g(x), x \geq 0$
ℓ_p	$x^p \ 0 < p < 1$
LSP	$\log(1+\frac{x}{\gamma})$
Laplace	$(1-\exp(-\frac{x}{\gamma}))$
LOG	$\log(\gamma+x)$
Logarithm	$\frac{1}{\log(\gamma+1)}\log(\gamma x+1)$
ETP	$\frac{1-\exp(-\gamma x)}{1-\exp(-\gamma)}$

II. NOTATIONS AND PRELIMINARIES

A. Notations

In this article, the fields of real and complex numbers are denoted as $\mathbb{R}$ and $\mathbb{C}$, respectively. Tensors are denoted by Euler script letters, e.g., $\mathcal{A}$; matrices are denoted by capital letters, e.g., A; sets are denoted by boldface capital letters, e.g., $\mathbf{A}$; vectors are denoted by boldface lowercase letters, e.g., $\boldsymbol{a}$, and scalars are denoted by lowercase letters, e.g., a.

More definitions and symbols are given as follows.

1) For $A \in \mathbb{R}^{n_1\times n_2}$, A^T is the transpose of A. $\text{rank}(A)$ and $\|A\|_*$ denote the rank function of matrix A (the number of nonzero singular values) and the nuclear norm of matrix A (the sum of singular values), respectively. $\sigma_i(A)$ denotes the ith largest singular value of matrix A, and $\sigma(A) = (\sigma_1(A), \sigma_2(A), \ldots, \sigma_r(A))^T$. $\cdot$ and $\otimes$ denote the matrix product and Kronecker product, respectively. I_n and F_n denote the $n \times n$ identity matrix and DFT matrix, respectively. $A \longrightarrow B$ indicates that B can be obtained by elementary row or column transformation of A.
2) A three-order tensor is denoted as $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$, where n_k $(k = 1, 2, 3)$ is a positive integer. Each element in this tensor is represented as $\mathcal{A}_{i_1i_2i_3}$. The Frobenius norm, l_1 norm, infinity norm, and l_0 norm are, respectively, denoted as $\|\mathcal{A}\|_F = (\sum_{i_1,i_2,i_3}\mathcal{A}^2_{i_1i_2i_3})^{1/2}$, $\|\mathcal{A}\|_1 = \sum_{i_1,i_2,i_3}|\mathcal{A}_{i_1i_2i_3}|$, $\|\mathcal{A}\|_\infty = \max_{i_1,i_2,i_3}|\mathcal{A}_{i_1i_2i_3}|$, and $\|\mathcal{A}\|_0$ (i.e., the number of nonzero entries of $\mathcal{A}$), respectively. For $\mathcal{A}$, $\mathcal{B} \in \mathbb{C}^{n_1\times n_2\times n_3}$, the inner product of $\mathcal{A}$ and $\mathcal{B}$ is denoted as $\langle\mathcal{A},\mathcal{B}\rangle = \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\mathcal{A}_{i_1i_2i_3}\text{Conj}(\mathcal{B}_{i_1i_2i_3})$. $\mathbf{0}$ denotes all-zero tensor.
3) $\#\mathbf{A}$ denotes the number of elements of set $\mathbf{A}$.

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4) Im(a), Re(a), and Conj(a) denote the imaginary part of a, real part of a, and conjugate of a, respectively.

Besides, following the definitions in [13] and [34], this article uses the MATLAB notation $\mathcal{A}(i,:,:)$, $\mathcal{A}(:,i,:)$, and $\mathcal{A}(:,:,i)$ to denote the ith horizontal, lateral, and frontal slice, respectively, and the frontal slice $\mathcal{A}(:,:,i)$ is denoted compactly as $A^{(i)}$. We use the MATLAB command fft to denote the result of DFT on $\mathcal{A}$ along the 3rd dimention, i.e., $\bar{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$. And we can get $\mathcal{A}$ from $\bar{\mathcal{A}}$ by $\mathcal{A} = \text{ifft}(\bar{\mathcal{A}}, [], 3)$, where the MATLAB command ifft is the inverse FFT. In addition, unfold($\cdot$), fold($\cdot$), bcirc($\cdot$), and bdiag($\cdot$) are defined as

$$\text{unfold}(\mathcal{A}) = \begin{pmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n_3)} \end{pmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2}$$

$$\text{fold}(\text{unfold}(\mathcal{A})) = \mathcal{A}$$

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

$$\text{bdiag}(\mathcal{A}) = \begin{pmatrix} A^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A^{(n_3)} \end{pmatrix}.$$

Note that $\text{bdiag}(\bar{\mathcal{A}}) = (F_{n_3} \otimes I_{n_1}) \cdot \text{bcirc}(\mathcal{A}) \cdot (F_{n_3}^{-1} \otimes I_{n_2})$, where $F_{n_3}^{-1}$ is the inverse of F_{n_3}.

B. Preliminary Definitions and Results

Definition 1 (t-Product [23]): Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3}$. Then, the t-product $\mathcal{A} *_t \mathcal{B}$ is defined as a tensor of size $n_1 \times l \times n_3$

$$\mathcal{A} *_t \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})). \tag{1}$$

Definition 2 (F-Diagonal Tensor [23]): Tensor $\mathcal{A}$ is called f-diagonal if each of its frontal slice is a diagonal matrix.

Definition 3 (Identity Tensor [23]): Tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is a tensor in which the first frontal slice is an identity matrix, and other frontal slices are all zeros.

Definition 4 (Mode-1 Conjugate Transpose): The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is denoted as $\mathcal{A}^{T_1} \in \mathbb{C}^{n_1 \times n_3 \times n_2}$, which is obtained by conjugate transposing each of the horizontal slice.

Definition 5 (Mode-2 Conjugate Transpose): The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is denoted as $\mathcal{A}^{T_2} \in \mathbb{C}^{n_3 \times n_2 \times n_1}$, which is obtained by conjugate transposing each of the lateral slice.

Definition 6 (Mode-3 Conjugate Transpose): The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is denoted as $\mathcal{A}^{T_3} \in \mathbb{C}^{n_2 \times n_1 \times n_3}$, which is obtained by conjugate transposing each of the frontal slice.

Definition 7 (Conjugate Transpose [12]): The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is denoted as $\mathcal{A}^* \in \mathbb{C}^{n_2 \times n_1 \times n_3}$, which is obtained by conjugate transposing each of the frontal slice and then reversing the order of the transposed frontal slice from position 2 to n_3.

Definition 8 (Orthogonal Tensor [23]): A tensor $\mathcal{Q} \in \mathbb{C}^{n \times n \times n_3}$ is orthogonal if it satisfies $\mathcal{Q}^* *_t \mathcal{Q} = \mathcal{Q} *_t \mathcal{Q}^* = \mathcal{I}$.

Theorem 1 (t-SVD [12]): Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then it can be factorized as $\mathcal{A} = \mathcal{U} *_t \mathcal{S} *_t \mathcal{V}^*$, 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, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor.

As stated in [12], we can get t-SVD of a tensor by Algorithm 1.

Algorithm 1 t-SVD [12]

Input: $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\lambda > 0$.
Output: $\mathcal{U}$, $\mathcal{S}$ and $\mathcal{V}$.
1. Compute $\bar{\mathcal{Y}} = \text{fft}(\mathcal{Y}, [], 3)$.
2. Compute each frontal slice of $\bar{\mathcal{U}}$, $\bar{\mathcal{S}}$ and $\bar{\mathcal{V}}$ from $\bar{\mathcal{Y}}$ by
for $i = 1, \ldots, \lfloor \frac{n_3+1}{2} \rfloor$ **do**
$[\bar{U}^{(i)}, \bar{S}^{(i)}, \bar{V}^{(i)}] = \text{SVD}(\bar{\mathcal{Y}}^{(i)})$;
end for
for $i = \lfloor \frac{n_3+1}{2} \rfloor + 1, \ldots, n_3$ **do**
$\bar{U}^{(i)} = Conj(\bar{U}^{(n_3-i+2)})$;
$\bar{S}^{(i)} = \bar{S}^{(n_3-i+2)}$;
$\bar{V}^{(i)} = Conj(\bar{V}^{(n_3-i+2)})$;
end for
3. $\mathcal{U} = \text{ifft}(\bar{U}, [], 3)$, $\mathcal{S} = \text{ifft}(\bar{S}, [], 3)$ and $\mathcal{V} = \text{ifft}(\bar{V}, [], 3)$.

Definition 9 (Tensor Tubal Rank [12]): For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor tubal rank, denoted as $\text{rank}_t(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ is obtained from the t-SVD of $\mathcal{A} = \mathcal{U} *_t \mathcal{S} *_t \mathcal{V}^*$. We can write $\text{rank}_t(\mathcal{A}) = \#\{i | \mathcal{S}(i,i,:) \neq \mathbf{0}\} = \#\{i | \mathcal{S}(i,i,1) \neq 0\}$. Denote $\sigma(\mathcal{S}) = (\mathcal{S}(1,1,1), \mathcal{S}(2,2,1), \ldots, \mathcal{S}(r,r,1))^T$, in which $r = \text{rank}_t(\mathcal{A})$.

Definition 10 (Tensor Nuclear Norm [12]): Let $\mathcal{A} = \mathcal{U} *_t \mathcal{S} *_t \mathcal{V}^*$ be the t-SVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The tensor nuclear norm of $\mathcal{A}$ is defined as $\|\mathcal{A}\|_* = \langle \mathcal{S}, \mathcal{I} \rangle = \sum_{i=1}^{r} \mathcal{S}(i,i,1)$, where $r = \text{rank}_t(\mathcal{A})$.

Definition 11 (Tensor Average Rank [12]): For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor average rank, denoted as $\text{rank}_a(\mathcal{A})$, is defined as $\text{rank}_a(\mathcal{A}) = (1/n_3)\text{rank}(\text{bcirc}(\mathcal{A}))$.

Definition 12 (Tensor Average Nuclear Norm [12]): For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor average nuclear norm of $\mathcal{A}$ is defined as $\|\mathcal{A}\|_{*,a} = (1/n_3)\|\text{bcirc}(\mathcal{A})\|_*$.

III. New Tensor Rank

In this section, the TVTR is discussed in detail, and a new tensor rank is given to better explore the low-rank structure within a data tensor.

A. Motivation: Transpose Variability of Tensor Recovery

It can be seen from Definitions 9 and 10 that the tensor tubal rank and tensor nuclear norm are base on t-SVD, in which discrete Fourier transform is applied on the 3rd dimension of the tensor. Therefore, the transpose operations of the tensor directly affects the tensor recovery methods

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II

TOP ROW: THE SINGULAR VALUES VECTORS OF THE SIX TENSORS. BOTTOM ROW: THE SINGULAR VALUES VECTORS OF THE BLOCK CIRCULANT MATRICES INCLUDING $\mathrm{bcirc}(\mathcal{A})$, $\mathrm{bcirc}(\mathcal{A}^{T_2})$, $\mathrm{bcirc}((\mathcal{A}^{T_2})^{T_3})$, $\mathrm{bcirc}(\mathcal{A}^{T_1})$, $\mathrm{bcirc}((\mathcal{A}^{T_1})^{T_3})$ AND $\mathrm{bcirc}(\mathcal{A}^{T_3})$.

	$\mathcal{A}$	$\mathcal{A}^{T_2}$	$(\mathcal{A}^{T_2})^{T_3}$	$\mathcal{A}^{T_1}$	$(\mathcal{A}^{T_1})^{T_3}$	$\mathcal{A}^{T_3}$
$\sigma(\cdot)$	$\begin{pmatrix}3.2988\\0.4407\end{pmatrix}$	$\begin{pmatrix}3.1631\\0.9183\end{pmatrix}$	$\begin{pmatrix}3.1631\\0.9183\end{pmatrix}$	$\begin{pmatrix}3.0859\\1.2910\end{pmatrix}$	$\begin{pmatrix}3.0859\\1.2910\end{pmatrix}$	$\begin{pmatrix}3.2988\\0.4407\end{pmatrix}$
$\sigma(\mathrm{bcirc}(\cdot))$	$\begin{pmatrix}3.7558\\2.8417\\0.5970\\0.2843\end{pmatrix}$	$\begin{pmatrix}3.7505\\2.5758\\1.2586\\0.5780\end{pmatrix}$	$\begin{pmatrix}3.7505\\2.5758\\1.2586\\0.5780\end{pmatrix}$	$\begin{pmatrix}3.3331\\2.8387\\1.5338\\1.0482\end{pmatrix}$	$\begin{pmatrix}3.3331\\2.8387\\1.5338\\1.0482\end{pmatrix}$	$\begin{pmatrix}3.7558\\2.8417\\0.5970\\0.2843\end{pmatrix}$

based on the two norms (including the tensor tubal rank and tensor nuclear norm). An example is given in the following to illustrate this point: Let $\mathcal{A} \in \mathbb{R}^{2\times2\times2}$, in which $\mathcal{A}^{(1)} = \begin{pmatrix}-0.1241 & 1.4090\\ 1.4897 & 1.4172\end{pmatrix}$, $\mathcal{A}^{(2)} = \begin{pmatrix}0.6715 & 0.7172\\ -1.2075 & 1.6302\end{pmatrix}$. For a three-order tensor $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, six tensors are obtained by all possible transpose operations for $\mathcal{A}$: $\mathcal{A} = (\mathcal{A}^{T_1})^{T_1} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{B}_1 = \mathcal{A}^{T_2} \in \mathbb{R}^{n_3\times n_2\times n_1}$, $\mathcal{B}_2 = (\mathcal{A}^{T_2})^{T_3} \in \mathbb{R}^{n_2\times n_3\times n_1}$, $\mathcal{B}_3 = \mathcal{A}^{T_1} \in \mathbb{R}^{n_1\times n_3\times n_2}$, $\mathcal{B}_4 = (\mathcal{A}^{T_1})^{T_3} \in \mathbb{R}^{n_3\times n_1\times n_2}$, and $\mathcal{B}_5 = \mathcal{A}^{T_3} \in \mathbb{R}^{n_2\times n_1\times n_3}$. From the top row of Table II, it can be seen that the tensor singular values of $\mathcal{A}$, $\mathcal{A}^{T_2}$, and $\mathcal{A}^{T_1}$ are different. Considering the following key optimal problem in low-rank tensor recovery:

$$\mathcal{D}(\mathcal{Y},\lambda) = \arg\min_{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}} \lambda\|\mathcal{X}\|_* + \frac{1}{2}\|\mathcal{Y}-\mathcal{X}\|_F^2 \quad (2)$$

this is the proximal operator of the tensor nuclear norm. To solve the problem shown in (2), Liu *et al.* [13] proposed an optimal algorithm. The whole algorithm is similar to Algorithm 2 given in Section V in this article. The only difference between the two algorithms is that, in Lu's work, a soft threshold with parameter λ is used instead of $\mathcal{T}_g(\bar{S}^{(i)},\lambda)$.[1] Therefore, the Lu's work is not considered in this article because of space limits. Algorithm 2 reveals that $\mathcal{D}(\mathcal{B}_i,\lambda)$ is not equivalent to the transpose of $\mathcal{D}(\mathcal{A},\lambda)$ for some λ, when $\sigma(\mathcal{A}) \neq \sigma(\mathcal{B}_i)$. Therefore, it can be concluded that the tensor nuclear norm based-tensor recovery methods have TVTR property. Note that, as stated in [12], (2) is equivalent to

$$\mathcal{D}(\mathcal{Y},\lambda) = \arg\min_{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}} \lambda\|\mathcal{X}\|_{*,a} + \frac{1}{2}\|\mathcal{Y}-\mathcal{X}\|_F^2. \quad (3)$$

Therefore, for the tensor average nuclear norm-based-tensor recovery methods, a similar conclusion can be obtained.

As discussed above, the effectiveness of the tensor recovery methods based on the two norms (including tensor nuclear norm and tensor average nuclear norm) is affected by the transpose operations on the data tensor, but this is ignored by traditional tensor recovery methods. An intuitive approach to solve this problem is to consider all possible transpose operations in the definition of tensor rank.

B. Weighted Tensor Average Rank

Definition 13 (Weighted Tensor Tubal Rank): Define weighted tensor tubal rank $\mathrm{rank}_{\mathrm{wt}}(\cdot)$ as

$$\mathrm{rank}_{\mathrm{wt}}(\mathcal{A}) = \sum_{k=1}^{3} \alpha_k \mathrm{rank}_t(\mathcal{A}^{T_k}) \quad (4)$$

where $\alpha_k (k = 1, 2, 3)$ indicates the weights which sum to 1.

Definition 14 (WTAR): Define weighted average tensor rank $\mathrm{rank}_{\mathrm{wa}}(\cdot)$ as

$$\mathrm{rank}_{\mathrm{wa}}(\mathcal{A}) = \sum_{k=1}^{3} \alpha_k \mathrm{rank}_a(\mathcal{A}^{T_k}) \quad (5)$$

where $\alpha_k (k = 1, 2, 3)$ indicates the weights which sum to 1.

Definition 15 (Weighted Tensor Nuclear Norm): Define weighted tensor nuclear norm $\|\cdot\|_{*,\mathrm{wt}}$ as

$$\|\mathcal{A}\|_{*,\mathrm{wt}} = \sum_{k=1}^{3} \alpha_k \|\mathcal{A}^{T_k}\|_* \quad (6)$$

where $\alpha_k (k = 1, 2, 3)$ indicates the weights which sum to 1.

Definition 16 (Weighted Tensor Average Nuclear Norm): Define weighted tensor average nuclear norm $\|\cdot\|_{*,\mathrm{wa}}$ as

$$\|\mathcal{A}\|_{*,\mathrm{wa}} = \sum_{k=1}^{3} \alpha_k \|\mathcal{A}^{T_k}\|_{*,a} \quad (7)$$

where $\alpha_k (k = 1, 2, 3)$ indicates the weights which sum to 1.

Property 1: For $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\sigma(\mathrm{bcirc}(\mathcal{A})) = \sigma(\mathrm{bcirc}(\mathcal{A}^{T_3}))$.

Proof:

$$\begin{aligned}
\mathrm{bcirc}(\mathcal{A}^{T_3}) &= \begin{pmatrix} A^{(1)T} & A^{(n_3)T} & \cdots & A^{(2)T} \\ A^{(2)T} & A^{(1)T} & \cdots & A^{(3)T} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)T} & A^{(n_3-1)T} & \cdots & A^{(1)T} \end{pmatrix} \\
&\longrightarrow \begin{pmatrix} A^{(1)T} & A^{(2)T} & \cdots & A^{(n_3)T} \\ A^{(2)T} & A^{(3)T} & \cdots & A^{(1)T} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)T} & A^{(1)T} & \cdots & A^{(n_3-1)T} \end{pmatrix} \\
&= \begin{pmatrix} A^{(1)} & A^{(2)} & \cdots & A^{(n_3)} \\ A^{(2)} & A^{(3)} & \cdots & A^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)} & A^{(1)} & \cdots & A^{(n_3-1)} \end{pmatrix}^T \\
&\longrightarrow \begin{pmatrix} A^{(1)} & A^{(n_3)} & \cdots & A^{(2)} \\ A^{(2)} & A^{(1)} & \cdots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)} & A^{(n_3-1)} & \cdots & A^{(1)} \end{pmatrix}^T \\
&= \mathrm{bcirc}(\mathcal{A})^T. \quad (8)
\end{aligned}$$

[1] $\mathcal{T}_g(Y,\lambda) = \arg\min_{X\in\mathbb{R}^{m\times n}} (1/2)\|Y-X\|_F^2 + \lambda\sum_{i=1}^{m}\sum_{j=1}^{n} g(|X_{ij}|)$.

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Therefore, Property 1 holds. ■

Theorem 2: For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, if $\alpha_1 = \alpha_2 = \alpha_3 = (1/3)$, $\|\mathcal{A}\|_{*,\mathrm{wa}} = \|\mathcal{A}^{T_s}\|_{*,\mathrm{wa}}$ for $s = 1, 2, 3$.

Proof: For $s = 1$, since $(\mathcal{A}^{T_1})^{T_1} = \mathcal{A}$ and $(\mathcal{A}^{T_1})^{T_2} = (\mathcal{A}^{T_2})^{T_3}$, $\|(\mathcal{A}^{T_1})^{T_1}\|_{*,a} = \|\mathcal{A}\|_{*,a} = \|\mathcal{A}^{T_3}\|_{*,a}$, $\|(\mathcal{A}^{T_1})^{T_2}\|_{*,a} = \|(\mathcal{A}^{T_2})^{T_3}\|_{*,a} = \|\mathcal{A}^{T_2}\|_{*,a}$, and $\|(\mathcal{A}^{T_1})^{T_3}\|_{*,a} = \|\mathcal{A}^{T_1}\|_{*,a}$ by Property 1.

Therefore, $\|\mathcal{A}^{T_1}\|_{*,\mathrm{wa}} = \sum_{k=1}^{3}(1/3)\|(\mathcal{A}^{T_1})^{T_k}\|_{*,a} = ((\|(\mathcal{A}^{T_1})^{T_1}\|_{*,a} + \|(\mathcal{A}^{T_1})^{T_2}\|_{*,a} + \|(\mathcal{A}^{T_1})^{T_3}\|_{*,a})/3) = \sum_{k=1}^{3}(1/3)\|\mathcal{A}^{T_k}\|_{*,a} = \|\mathcal{A}\|_{*,\mathrm{wa}}$.

For $s = 2$, since $(\mathcal{A}^{T_2})^{T_1} = (\mathcal{A}^{T_1})^{T_3}$ and $(\mathcal{A}^{T_2})^{T_2} = \mathcal{A}$, $\|(\mathcal{A}^{T_2})^{T_1}\|_{*,a} = \|(\mathcal{A}^{T_1})^{T_3}\|_{*,a} = \|\mathcal{A}^{T_1}\|_{*,a}$, $\|(\mathcal{A}^{T_2})^{T_2}\|_{*,a} = \|\mathcal{A}\|_{*,a} = \|\mathcal{A}^{T_3}\|_{*,a}$, and $\|(\mathcal{A}^{T_2})^{T_3}\|_{*,a} = \|\mathcal{A}^{T_2}\|_{*,a}$ by Property 1.

Therefore, $\|\mathcal{A}^{T_2}\|_{*,\mathrm{wa}} = \sum_{k=1}^{3}(1/3)\|(\mathcal{A}^{T_2})^{T_k}\|_{*,a} = ((\|(\mathcal{A}^{T_2})^{T_1}\|_{*,a} + \|(\mathcal{A}^{T_2})^{T_2}\|_{*,a} + \|(\mathcal{A}^{T_2})^{T_3}\|_{*,a})/3) = \sum_{k=1}^{3}(1/3)\|\mathcal{A}^{T_k}\|_{*,a}$.

For $s = 3$, since $(\mathcal{A}^{T_3})^{T_1} = (\mathcal{A}^{T_2})^{T_3}$ and $(\mathcal{A}^{T_3})^{T_2} = (\mathcal{A}^{T_1})^{T_3}$, $\|\mathcal{A}^{T_3}\|_{*,\mathrm{wa}} = \sum_{k=1}^{3}(1/3)\|(\mathcal{A}^{T_3})^{T_k}\|_{*,a} = ((\|(\mathcal{A}^{T_3})^{T_1}\|_{*,a} + \|(\mathcal{A}^{T_3})^{T_2}\|_{*,a} + \|(\mathcal{A}^{T_3})^{T_3}\|_{*,a})/3) = ((\|(\mathcal{A}^{T_2})^{T_3}\|_{*,a} + \|(\mathcal{A}^{T_1})^{T_3}\|_{*,a} + \|(\mathcal{A}^{T_3})^{T_3}\|_{*,a})/3) = \sum_{k=1}^{3}(1/3)\|\mathcal{A}^{T_k}\|_{*,a}$. ■

Since $\|\mathcal{A}\|_* = \|\mathcal{A}\|_{*,a}$ as stated in [12], we have $\|\mathcal{A}\|_{*,\mathrm{wa}} = \|\mathcal{A}\|_{*,\mathrm{wt}}$. Therefore, the following theorem is derived.

Theorem 3: For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, if $\alpha_1 = \alpha_2 = \alpha_3 = (1/3)$, $\|\mathcal{A}\|_{*,\mathrm{wt}} = \|\mathcal{A}^{T_s}\|_{*,\mathrm{wt}}$ for $s = 1, 2, 3$.

IV. TRPCA WITH WEIGHTED TENSOR AVERAGE RANK

Based on the definition of $\mathrm{rank}_{\mathrm{wt}}(\cdot)$, TPRCA with WTAR is defined as follows:

$$\min_{\mathcal{L},\mathcal{S}} \ \mathrm{rank}_{\mathrm{wa}}(\mathcal{L}) + \lambda\|\mathcal{S}\|_0 \quad \text{s.t. } \|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F \le \delta \tag{9}$$

where $\mathcal{P} = \mathcal{L} + \mathcal{S} + \mathcal{Z}$; $\mathcal{L}$ is low-rank; $\mathcal{S}$ is sparse, and $\mathcal{Z}$ is a small noisy perturbation and $\|\mathcal{Z}\|_F \le \delta$. Since $\mathrm{rank}_{\mathrm{wt}}(\cdot)$ and ℓ_0 norm is discrete, the continuous version of (9) is considered, which is defined as follows:

$$\min_{\mathcal{L},\mathcal{S}} \ \|\mathcal{L}\|_{*,\mathrm{wa}}^{g} + \lambda\|\mathcal{S}\|_g \quad \text{s.t. } \|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F \le \delta \tag{10}$$

where $\|\mathcal{L}\|_{*,\mathrm{wa}}^{g} = \sum_{k=1}^{3}(\alpha_k/n_k)\sum_{i=1}^{r_k} g(\sigma_i(\mathrm{bcirc}(\mathcal{L}^{T_k})))$, $\|\mathcal{S}\|_g = \sum_{i_1,i_2,i_3} g(|\mathcal{S}_{i_1 i_2 i_3}|)$, and $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is an increasing function. Note that all the surrogate functions of ℓ_0 listed in Table I satisfy this condition.

Remark 1: From Property 1, we can get the same conclusion with Theorem 2 easily for $\|\cdot\|_{*,\mathrm{wa}}^{g}$, i.e., if $\alpha_1 = \alpha_2 = \alpha_3 = (1/3)$, $\|\mathcal{A}\|_{*,\mathrm{wa}}^{g} = \|\mathcal{A}^{T_s}\|_{*,\mathrm{wa}}^{g}$ for $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $s = 1, 2, 3$.

A. ℓ_p Minimization Formulation

Taking $g(\cdot)$ in (10) as ℓ_p norm, then (10) is turned to

$$\min_{\mathcal{L},\mathcal{S}} \ \|\mathcal{L}\|_{p,\mathrm{wa}}^{p} + \lambda\|\mathcal{S}\|_{p,p}^{p} \quad \text{s.t. } \|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F \le \delta \tag{11}$$

where $\|\mathcal{L}\|_{p,\mathrm{wa}}^{p} = \sum_{k=1}^{3}(\alpha_k/n_k)(\sum_{i=1}^{r_k}\sigma_i(\mathrm{bcirc}(\mathcal{L}^{T_k}))^{(1/p)})^p$, $r_k = \mathrm{rank}(\mathrm{bcirc}(\mathcal{L}^{T_k}))$, and $\|\mathcal{S}\|_{p,p}^{p} = (\sum_{i_1,i_2,i_3}|\mathcal{S}_{i_1 i_2 i_3}|^{(1/p)})^p$. For convenience, (11) is referred to as TRWTAR-ℓ_p (where TRWTAR and ℓ_p stand for Tensor Recovery with WTAR and ℓ_p norm, respectively). It is easy to see that, for $p = 1$, (11) reduces to

$$\min_{\mathcal{L},\mathcal{S}} \ \|\mathcal{L}\|_{*,\mathrm{wa}} + \lambda\|\mathcal{S}\|_1 \quad \text{s.t. } \|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F \le \delta \tag{12}$$

which is referred to as tensor recovery with weighted tensor average nuclear norm (TRWTANN).

B. Worst Case Error Bound

Here, an error bound is established under the transformed ℓ_p minimization problem (11).

Theorem 4: Let $(\mathcal{L}_0, \mathcal{S}_0)$ be the pair of true low-rank and sparse tensors, and $\mathcal{L}^*$ be the solution to the optimization problem (11). If the average of the entries of the sparse component $\mathcal{S}_0$ is bounded by T, and the carnality of the support $\mathcal{S}_0$ is bounded by m, then $\mathrm{Err}(\mathcal{L}^*) = ((\|\mathcal{L}_0 - \mathcal{L}^*\|_F)/M) \le \sqrt[p]{((2mT^p + ((2\delta)^p/M^{(p/2)-1}))/(M^p(1 - (1/\lambda))))}$, where $\lambda > 1$, $M = \prod_{k=1}^{3} n_k$. Remark $n^{(1)} = n_2$, $n^{(2)} = n_1$ and $n^{(3)} = n_3$.

The proof of Theorem 4 is given in the appendix.

To give an intuitive understanding of Theorem 4, consider the two most simple cases.

1) For $p = 1$, $\delta = 0$, we have

$$\mathrm{Err}(\mathcal{L}^*) = \frac{2mT}{M\left(1 - \frac{1}{\lambda}\right)}$$

where $(m/M) \ll 1$ is the sparsity coefficient, and T is bounded. Usually, the entries in visual data are typically bounded by a constant that is not too large, i.e., the biggest value of entry is 255 for images. Thus, the error bound is rather small, indicating rather good recovery.

2) For $p = 1$, $T = 0$, we have

$$\mathrm{Err}(\mathcal{L}^*) = \frac{2\delta}{M^{\frac{1}{2}}\left(1 - \frac{1}{\lambda}\right)}$$

where $(1/(M^{(1/2)}(1 - (1/\lambda)))) \ll 1$ for $\lambda = \infty$. As suggested in the above, λ in (12) should be set to a large enough value for $\mathcal{S}_0 = \mathbf{0}$.

V. GENERAL ALGORITHM

This section introduces a general optimization algorithm for solving (10).

A. Key Problem

To solve problem (10), the following subproblem is considered:

$$\arg\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \ \lambda\|\mathcal{X}\|_{*,a}^{g} + \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F^2 \tag{13}$$

where $\|\mathcal{X}\|_{*,a}^{g} = (1/n_3)\sum_{i=1}^{r} g(\sigma_i(\mathrm{bcirc}(\mathcal{X})))$, and $r = \mathrm{rank}(\mathrm{bcirc}(\mathcal{X}))$. In the following, we will prove that (13) can be solved by GTSVT (Algorithm 2), where

$$\mathcal{T}_g(Y, \lambda) = \arg\min_{X \in \mathbb{R}^{m \times n}} \ \frac{1}{2}\|Y - X\|_F^2 + \lambda\sum_{i=1}^{m}\sum_{j=1}^{n} g(|X_{ij}|).$$

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Algorithm 2 Generalized Tensor Singular Value Thresholding (GTSVT)

Input: $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\lambda > 0$.
Output: $\mathcal{D}_g(\mathcal{Y}, \lambda)$.
1. Compute $\bar{\mathcal{Y}} = \text{fft}(\mathcal{Y}, [], 3)$.
2. Perform the generalized SVT [41] on each frontal slice of $\bar{\mathcal{Y}}$ by
for $i = 1, \ldots, \lfloor \frac{n_3+1}{2} \rfloor$ **do**
$[\bar{U}^{(i)}, \bar{S}^{(i)}, \bar{V}^{(i)}] = \text{SVD}(\bar{\mathcal{Y}}^{(i)})$;
$\bar{W}^{(i)} = \bar{U}^{(i)} \mathcal{T}_g(\bar{S}^{(i)}, \lambda) \bar{V}^{(i)*}$;
end for
for $i = \lfloor \frac{n_3+1}{2} \rfloor + 1, \ldots, n_3$ **do**
$\bar{W}^{(i)} = Conj(\bar{W}^{(n_3-i+2)})$;
end for
3. $\mathcal{D}_g(\mathcal{Y}, \lambda) = \text{ifft}(\bar{\mathcal{W}}, [], 3)$.

Theorem 5: For any $\lambda > 0$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, if g is increasing on $[0, +\infty)$, then the tensor singular value thresholding operator obeys

$$\mathcal{D}_g(\mathcal{Y}, \lambda) \in \arg \min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \lambda \|\mathcal{X}\|_{*,a}^g + \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F^2. \tag{14}$$

The proof of Theorem 5 is given in the appendix.

B. General Algorithm Based on Alternating Direction Method

Equation (10) can be reduced to

$$\min_{\mathcal{L}, \mathcal{S}} \alpha \sum_{k=1}^{3} \alpha_k \|\mathcal{L}^{T_k}\|_{*,a}^g + \beta \|\mathcal{S}\|_g + \frac{1}{2}\|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F^2 \tag{15}$$

where $\|\mathcal{L}^{T_k}\|_{*,a}^g = (1/n_k)\sum_{i=1}^{r_k} g(\sigma_i(\text{bcirc}(\mathcal{L}^{T_k})))$. To simplify (15), a series of auxiliary tensors $\mathcal{M}_k (k = 1, 2, 3)$ are introduced to replace $\mathcal{L}^{T_k}$ and to remove the correlation of $\mathcal{L}^{T_k}$. Then, (15) can be rewritten to

$$\begin{aligned} &\min_{\mathcal{M}_k, \mathcal{L}, \mathcal{S}} \frac{1}{2}\|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F^2 + \alpha \sum_{k=1}^{3} \alpha_k \|\mathcal{M}_k\|_{*,a}^g + \beta \|\mathcal{S}\|_g \\ &\quad \text{s.t. } \mathcal{L}^{T_k} = \mathcal{M}_k, \quad k = 1, 2, 3. \end{aligned} \tag{16}$$

To relax the above equality constraints, this article applies the alternating direction method (ADMM) [34] to the above problem, and the following augmented Lagrangian function is obtained:

$$\begin{aligned} &f_\mu(\mathcal{M}_k, \mathcal{L}, \mathcal{S}, \mathcal{Q}_k) \\ &= \frac{1}{2}\|\mathcal{P} - \mathcal{L} - \mathcal{S}\|_F^2 + \alpha \sum_{k=1}^{3} \alpha_k \|\mathcal{M}_k\|_{*,a}^g + \beta \|\mathcal{S}\|_g \\ &\quad + \Sigma_{k=1}^3 \left(\langle \mathcal{Q}_k, \mathcal{L}^{T_k} - \mathcal{M}_k \rangle + \frac{\mu_k}{2}\|\mathcal{L}^{T_k} - \mathcal{M}_k\|_F^2 \right) \end{aligned} \tag{17}$$

where μ_i is a positive scalar, and $\mathcal{Q}_k$ is Lagrange multiplier tensor. According to the framework of ADMM, the above optimization problem can be iteratively solved as follows.

Step 1: Given $\mathcal{L}^{(s)}$ and $\mathcal{Q}_k^{(s)}$, update $\mathcal{M}_k, k = 1, 2, 3$ by

$$\begin{aligned} \mathcal{M}_k^{(s+1)} &= \arg \min_{\mathcal{M}_k} \frac{\mu_k}{2} \left\| (\mathcal{L}^{(s)})^{T_k} - \mathcal{M}_k + \frac{1}{\mu_k}\mathcal{Q}_k^{(s)} \right\|_F^2 \\ &\quad + \alpha \alpha_k \|\mathcal{M}_k\|_{*,a}^g \\ &= \mathcal{D}_g\left((\mathcal{L}^{(s)})^{T_k} + \frac{1}{\mu_k}\mathcal{Q}_k^{(s)}, \frac{\alpha \alpha_k}{\mu_k} \right). \end{aligned} \tag{18}$$

Step 2: Given $\mathcal{M}_k^{(s+1)}$, $\mathcal{S}^{(s)}$ and $\mathcal{Q}_k^{(s)}, k = 1, 2, 3$, update $\mathcal{L}$ by

$$\begin{aligned} \mathcal{L}^{(s+1)} &= \arg \min_{\mathcal{L}} \frac{1}{2}\|\mathcal{P} - \mathcal{L} - \mathcal{S}^{(s)}\|_F^2 \\ &\quad + \sum_{k=1}^{3} \frac{\mu_k}{2} \left\| \mathcal{L}^{T_k} - \mathcal{M}_k^{(s+1)} + \frac{1}{\mu_k}\mathcal{Q}_k^{(s)} \right\|_F^2. \end{aligned} \tag{19}$$

Calculate the partial derivative of the above formulation with respect to $\mathcal{L}$, and set it to zero

$$-\mathcal{P} + \mathcal{L} + \mathcal{S}^{(s)} + \Sigma_{k=1}^3 \mu_k \left(\mathcal{L} - \left(\mathcal{M}_k^{(s+1)} - \frac{1}{\mu_k}\mathcal{Q}_k^{(s)} \right)^{T_k} \right) = 0.$$

By rearranging the term with $\mathcal{L}$, we have

$$\mathcal{L}^{(s+1)} = \frac{\mathcal{P} - \mathcal{S}^{(s)} + \Sigma_{k=1}^3 \mu_k \left(\mathcal{M}_k^{(s+1)} - \frac{1}{\mu_k}\mathcal{Q}_k^{(s)} \right)^{T_k}}{1 + \Sigma_{k=1}^3 \mu_k}. \tag{20}$$

Step 3: Given $\mathcal{L}^{(s+1)}$, update $\mathcal{S}$ by

$$\mathcal{S}^{(s+1)} = \arg \min_{\mathcal{S}} \frac{1}{2}\|\mathcal{P} - \mathcal{L}^{(s+1)} - \mathcal{S}\|_F^2 + \beta \|\mathcal{S}\|_g \tag{21}$$

$$= \mathcal{T}_g(\mathcal{P} - \mathcal{L}^{(s+1)}, \beta). \tag{22}$$

Step 4: Given $\mathcal{Q}_k^{(s)}$, $\mathcal{L}^{(s+1)}$ and $\mathcal{M}_k^{(s+1)}, k = 1, 2, 3$, update $\mathcal{L}$ by

$$\mathcal{Q}_k^{(s+1)} = \mathcal{Q}_k^{(s)} + \mu_k \left((\mathcal{L}^{(s+1)})^{T_k} - \mathcal{M}_k^{(s+1)} \right) \quad \forall k. \tag{23}$$

VI. Experimental Results

In this section, four sets of experiments are conducted to illustrate the effectiveness of our proposed methods. The first set of experiments are performed on the color image data contaminated by zero-mean Gaussian noise, and the proposed methods including TRWTANN and TRWTAR-ℓ_p are compared with several state-of-the-art low-rank tensor recovery methods, including SNN [11], Liu's work (called Liu for short in the following) [10], SRALT-ℓ_p [13], KBR [35], and TRPCA [12]. The second and third sets of experiments are performed on the color image data and video, respectively. All of them are contaminated by the mixture of zero-mean Gaussian noise and random valued impulse noise in different noise levels to test the seven methods. To illustrate the robustness of the proposed methods to outliers in the visual data and their effectiveness in practical applications, in the fourth set of experiments, all seven methods are tested on background subtraction. The source code of SRALT-ℓ_p[2] and KBR[3] are provided by their authors, while the source code of the remaining methods

[2]https://github.com/18357710774/SRALT_code
[3]https://github.com/XieQi2015/KBR-TC-and-RPCA

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
COLOR IMAGE DENOISING RESULTS (PSNR) BY DIFFERENT METHODS

	$\delta = 5$							$\delta = 10$						
	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
House	29.77	29.58	**36.06**	35.16	31.43	31.88	34.37	28.36	28.25	29.38	29.70	31.07	**31.52**	31.04
Peppers	32.02	31.76	**34.70**	32.21	31.91	32.77	34.55	28.29	28.29	28.66	29.49	30.81	**31.06**	30.97
Lena	32.71	32.57	**35.51**	33.80	32.69	33.53	35.32	28.50	28.56	29.17	30.29	31.46	31.74	**31.82**
Baboon	31.73	31.16	**34.12**	27.65	29.72	30.67	33.15	28.04	27.92	28.22	26.36	28.64	**29.10**	28.66
F16	33.39	33.25	**36.40**	33.27	33.53	34.42	36.39	28.71	28.78	30.15	30.02	31.97	32.12	**32.54**
Kodak image1	33.19	33.06	35.95	33.31	33.99	35.01	**37.24**	28.37	28.57	28.76	30.07	31.62	31.71	**32.35**
Kodak image2	32.68	32.61	36.55	33.90	34.76	35.71	**37.43**	28.29	28.53	28.75	30.75	32.50	32.42	**33.00**
Kodak image3	32.80	32.71	36.85	34.36	34.64	35.60	**37.41**	28.32	28.57	28.89	30.80	32.29	32.29	**32.92**
Kodak image12	33.04	32.99	37.22	36.14	35.51	36.43	**38.12**	28.39	28.65	29.87	31.58	32.95	32.78	**33.71**
Average	32.37	32.19	35.93	33.31	33.13	34.00	**36.00**	28.36	28.46	29.10	29.90	31.48	31.64	**31.89**
	$\delta = 15$							$\delta = 20$						
	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
House	22.98	22.96	**29.53**	27.11	27.14	26.50	28.94	22.46	22.49	**28.40**	24.78	26.95	26.56	26.59
Peppers	26.23	25.68	28.85	27.22	27.75	27.73	**28.96**	25.24	24.88	27.36	25.43	26.90	27.17	**27.47**
Lena	26.74	26.22	28.96	27.87	28.52	28.58	**29.87**	25.69	25.38	28.22	25.93	27.59	28.02	**28.31**
Baboon	23.30	22.34	25.72	24.73	25.03	24.95	**26.29**	22.65	21.86	**24.84**	23.52	24.54	24.65	24.48
F16	27.14	26.34	28.52	27.59	28.94	28.87	**30.31**	26.00	25.54	28.42	25.58	27.88	28.17	**28.58**
Kodak image1	26.47	26.13	28.35	27.66	28.58	28.71	**29.61**	25.02	25.13	26.53	25.79	27.22	27.54	**28.05**
Kodak image2	27.45	27.79	**30.77**	28.63	30.25	30.39	30.60	25.91	26.76	26.87	26.76	28.56	29.00	**29.71**
Kodak image3	27.16	27.40	30.36	28.41	29.75	29.94	**30.41**	25.61	26.32	27.35	26.44	28.14	28.61	**29.27**
Kodak image12	27.57	27.98	29.53	28.74	30.78	30.95	**31.23**	25.99	26.86	29.67	26.58	28.89	29.38	**30.35**
Average	26.12	25.87	28.96	27.55	28.53	28.51	**29.58**	24.95	25.02	27.52	25.65	27.41	27.68	**28.09**
	$\delta = 25$							$\delta = 30$						
	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
House	21.34	21.36	**25.24**	23.13	24.22	23.95	24.01	20.94	20.86	22.30	21.62	**24.19**	23.93	24.07
Peppers	24.06	23.74	24.20	23.98	25.04	25.40	**25.51**	23.39	23.14	21.37	22.82	24.63	25.04	**25.36**
Lena	24.66	24.34	25.32	24.27	25.87	26.25	**26.26**	23.98	23.70	22.22	22.90	25.46	25.96	**26.26**
Baboon	21.23	20.85	**22.92**	22.34	22.42	22.69	22.54	20.82	20.44	20.79	21.36	22.27	**22.56**	22.54
F16	24.90	24.45	**27.44**	24.01	26.13	26.34	26.38	24.14	23.84	24.67	22.72	25.65	26.02	**26.32**
Kodak image1	24.25	24.24	23.50	24.24	25.76	26.14	**26.15**	23.39	23.59	20.98	22.96	25.10	25.40	**25.76**
Kodak image2	25.65	26.15	23.20	25.23	27.87	28.20	**28.23**	24.68	25.41	20.63	23.92	26.91	27.17	**27.71**
Kodak image3	25.15	25.50	23.62	24.80	27.13	27.57	**27.71**	24.17	24.71	20.93	23.80	26.23	26.62	**27.07**
Kodak image12	25.73	26.23	27.70	24.88	28.24	28.60	**28.70**	24.70	25.44	23.96	23.48	27.14	27.43	**28.13**
Average	24.11	24.09	24.79	24.10	25.85	26.13	**26.17**	23.36	23.46	21.98	22.84	25.29	25.57	**25.91**

including SNN, Liu's work, and TRPCA are provided by the LibADMM toolbox.[4] The parameters of all methods are tuned to the best for each case. In addition, $\alpha_k (1 \leq k \leq 3)$ in our methods are set to (1/3).

A. Zero-Mean Gaussian Noise: Color Image Denoising

The clean color image with a size of $n_1 \times n_2 \times 3$ can be approximated by low-rank tensor $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times 3}$, and the zero-mean Gaussian noise can be regarded as small entry-wise perturbations $\mathcal{Z}_0 \in \mathbb{R}^{n_1 \times n_2 \times 3}$, which is a tensor with the entries independently sampled from a $\mathcal{N}(0, \delta^2)$ distribution (the noised image can be obtained by $\mathcal{P} = \mathcal{L}_0 + \mathcal{Z}_0$). In this part, all the seven methods (including SNN, Liu, SRALT-ℓ_p, KBR, TRPCA, TRWTANN, and TRWTAR-ℓ_p) are applied to color image recovery in which the color image is contaminated by zero-mean Gaussian noise. All methods are performed on House, Lena, Peppers, F16, Baboon, and the 1–3th and 12th images from the Kadak PhotoCD.[5] Meanwhile, the standard deviations of zero-mean Gaussian noise δ are set to $\delta = \{5, 10, 15, 20, 25, 30\}$.

Table III shows the peak signal-to-noise ratio (PSNR) results of different methods when the image data is corrupted by zero-mean Gaussian noise, and the highest PSNR values are marked in bold. The visual quality performance of all the methods is reported in Figs. 1 and 2. From these results, the following observations are made. First, the PSNR results of the proposed methods (TRWTANN and TRWTAR-ℓ_p) and other five methods (SNN, Liu, SRALT-ℓ_p, KBR, and TRPCA) indicate that TRWTANN and TRWTAR-ℓ_p achieve the best denoising performance in most cases. Especially, for case of $\delta = 15$, TRWTAR-ℓ_p even outperforms the five comparing methods by at least 1 dB on average PSNR. This illustrates the effectiveness of the methods based on the WTAR for handling Gaussian noise. Besides, the PSNR results of TRWTANN and TRWTAR-ℓ_p indicate that using the nonconvex surrogate strategy given in this article can improve the effectiveness of the original method (TRWTANN) significantly. In addition, from Figs. 1 and 2, it can be seen that the three tensor recovery methods based on t-product (including TRPCA, TRWTANN, and TRWTAR-ℓ_p) retain more information and details about image, while the denoised images obtained by SNN and Liu appear some white stripes. For the remaining two methods including SRALT-ℓ_p and KBR, there are still some residual noise within the denoised image. This validates the effectiveness of the methods based on t-product.

B. Zero-Mean Gaussian-Impulse Mixed Noise: Color Image Denoising

In this part, the proposed models are applied to image recovery, where the color image is contaminated by the mixture of zero-mean Gaussian noise $\mathcal{Z}_0$ and random valued impulse noise. Because the clean color image can be approximated by low-rank tensors, and the random valued impulse noise with density level c can be regarded as sparse errors $\mathcal{S}_0$,[6] the noise can be removed from the color images $\mathcal{P} = \mathcal{L}_0 + \mathcal{Z}_0 + \mathcal{S}_0$ by all the seven methods (including SNN, Liu, SRALT-ℓ_p, KBR, TRPCA, TRWTANN, and TRWTAR-ℓ_p). All the methods are tested on the testing

[4] https://github.com/canyilu/LibADMM-toolbox
[5] https://webpages.tuni.fi/foi/GCF-BM3D/index.html

[6] $c3n_1n_2$ entries in $\mathcal{S}_0$ uniformly distributed in $[0, 255]$, and the remain entries in $\mathcal{S}_0$ are zeros.

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

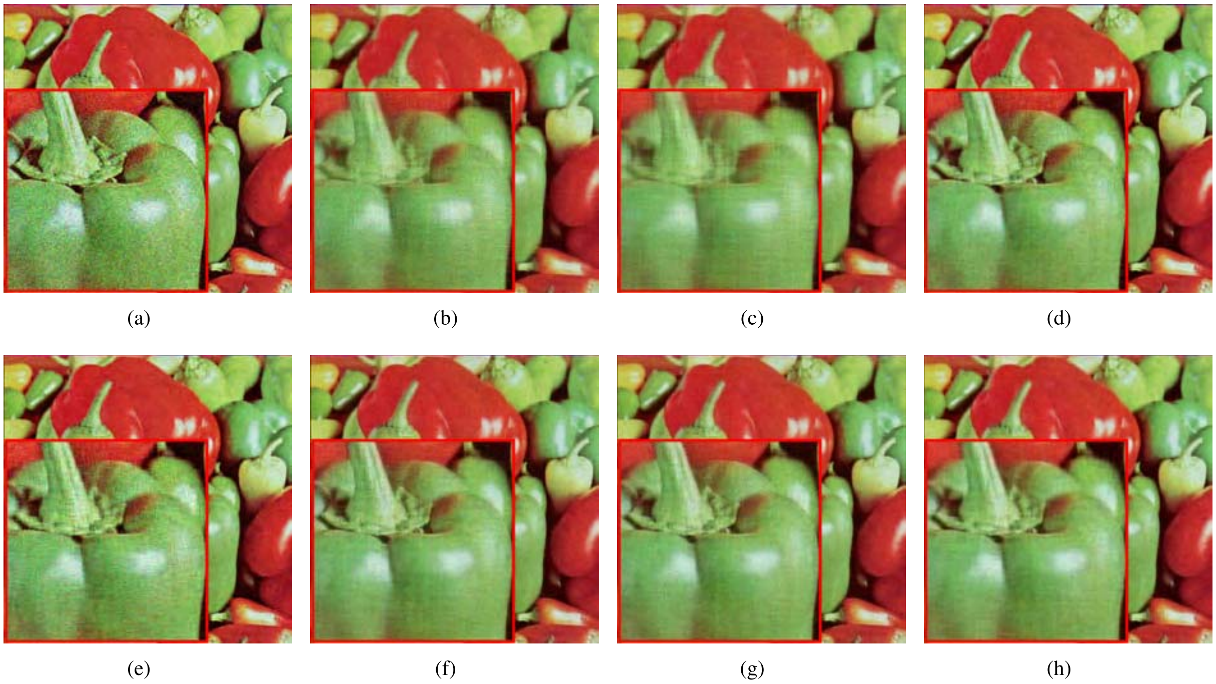

Fig. 1. Denoised results on "Peppers," $\delta = 20$. (a) Noised image. (b) SNN. (c) Liu. (d) SRALT-ℓ_p. (e) KBR. (f) TRPCA. (g) TRWTANN. (h) TRWTAR-ℓ_p.

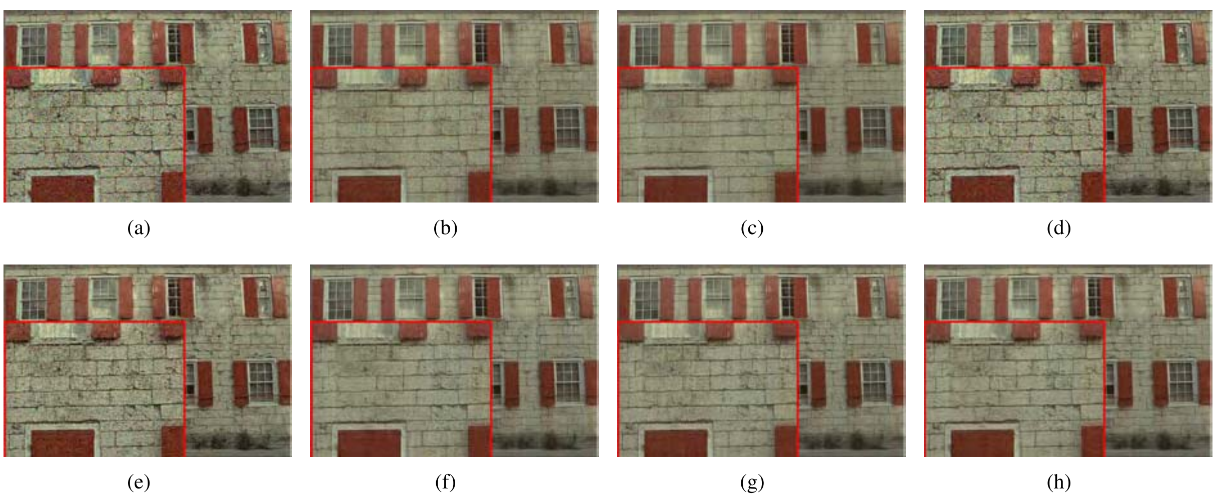

Fig. 2. Denoised results on "kodak image1," $\delta = 30$. (a) Noised image. (b) SNN. (c) Liu. (d) SRALT-ℓ_p. (e) KBR. (f) TRPCA. (g) TRWTANN. (h) TRWTAR-ℓ_p.

image set that contains House, Lena, Peppers, F16, Baboon, and the 1–3th and 12th images from the Kadak PhotoCD. Meanwhile, the noise is set to zero-mean Gaussian noise with standard deviations δ and random-valued impulse noise with density level c. Besides, in this experiment, (δ, c) is set to $(\delta, c) = \{(0, 5\%), (5, 5\%), (5, 10\%), (15, 10\%), (15, 15\%),$ and $(30, 15\%)\}$.

All the methods are evaluated by the PSNR value and visual results. From Table IV, it can be seen that the proposed methods (TRWTANN and TRWTAR-ℓ_p) outperform SNN, Liu, SRALT-ℓ_p, KBR, and TRPCA by a large margin in all cases on PSNR values. As shown in Figs. 3 and 4, the proposed methods retain more details in the denoised images. These results indicate the superiority of the proposed methods. The performance superiority is achieved by considering different tensor transpose operations in the progress of estimating the latent low-rank tensor, which makes use of the information within the tensor data as much as possible. This illustrates that the new tensor rank (WTAR) given in this article (see Definition 14) is more reasonable in real applications than others.

C. Zero-Mean Gaussian-Impulse Mixed Noise: Video Sequence Denoising

Similar to the case of color image denoising, video sequence denoising can also be regarded as a low-rank tensor recovery problem. In this case, each color frame of video is folded in the third dimension of the data tensor $\hat{\mathcal{L}}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times 3}$ (corresponding to the color video with size of $n_1 \times n_2 \times n_3 \times 3$) to obtain clean tensor data $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times 3}$. Then, TRWTANN and TRWTAR-ℓ_p are compared with the other five methods

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
COLOR IMAGE DENOISING RESULTS (PSNR) BY DIFFERENT METHODS

	$(\delta, c) = (0, 5\%)$							$(\delta, c) = (5, 5\%)$						
	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
Baboon	29.72	29.61	26.81	27.06	29.50	29.43	**29.99**	28.32	28.20	25.73	26.56	28.34	28.76	**29.34**
F16	35.84	36.03	34.66	33.91	35.71	36.40	**37.69**	31.81	31.79	32.19	32.09	32.22	33.17	**34.01**
House	30.36	31.15	32.78	**39.41**	31.53	32.63	32.77	29.73	30.19	30.99	**33.51**	30.65	30.85	30.71
Lena	34.66	34.73	35.56	35.61	34.88	35.48	**36.31**	31.37	31.27	32.47	32.80	31.87	32.82	**33.31**
Peppers	33.09	33.21	33.56	33.15	32.62	33.58	**34.40**	30.68	30.61	29.68	31.38	30.69	31.46	**31.92**
Kodak image1	34.23	35.07	30.15	33.15	**38.68**	37.11	38.61	30.49	30.82	27.67	31.99	32.15	33.20	**33.99**
Kodak image2	34.39	35.67	27.56	35.21	37.21	37.80	**39.49**	31.42	31.54	26.00	33.15	32.12	33.62	**34.25**
Kodak image3	35.37	35.69	30.35	35.87	35.99	37.40	**38.96**	31.41	31.55	27.71	33.38	31.91	33.45	**34.06**
Kodak image12	36.09	36.48	35.12	38.29	38.48	38.98	**40.46**	31.63	31.81	32.23	34.78	32.56	34.32	**34.85**
Average	33.86	34.48	31.84	34.63	34.96	35.42	**36.52**	30.76	30.86	29.41	32.18	31.39	32.41	**32.94**
	$(\delta, c) = (10, 10\%)$							$(\delta, c) = (15, 10\%)$						
	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
Baboon	23.23	23.16	20.48	24.46	22.74	23.95	**25.26**	22.67	22.57	23.04	23.15	22.21	23.41	**23.55**
F16	26.97	26.72	22.43	28.21	26.50	27.67	**28.89**	25.92	25.73	24.96	25.67	25.65	26.84	**26.85**
House	23.24	24.03	23.40	**26.87**	24.58	24.97	26.06	22.77	23.51	**24.56**	24.30	23.92	24.55	24.13
Lena	27.14	27.09	24.24	28.60	27.11	28.00	**29.04**	26.08	25.96	26.32	26.16	26.05	27.14	**27.22**
Peppers	26.48	26.34	24.25	27.81	25.80	26.88	**27.91**	25.49	25.36	**26.29**	25.60	24.94	26.06	26.11
Kodak image1	26.66	26.50	23.85	27.88	27.37	27.87	**28.78**	25.25	25.23	25.85	25.61	25.93	26.12	**26.65**
Kodak image2	28.21	28.24	27.23	29.22	28.37	29.34	**29.44**	26.56	26.74	26.02	26.88	27.05	27.49	**27.50**
Kodak image3	28.03	28.10	26.55	29.23	27.98	29.05	**29.30**	26.31	26.52	26.66	26.81	26.63	27.15	**27.30**
Kodak image12	28.33	28.38	25.71	29.60	28.89	**29.94**	29.92	26.61	26.80	27.61	26.93	27.40	27.79	**27.85**
Average	26.48	26.51	24.24	27.99	26.59	27.52	**28.29**	25.30	25.38	25.70	25.68	25.53	26.28	**26.35**
	$(\delta, c) = (15, 15\%)$							$(\delta, c) = (30, 15\%)$						
	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
Baboon	20.64	20.81	21.67	22.27	19.98	22.38	**22.41**	19.67	19.78	20.01	19.71	19.04	**20.12**	20.06
F16	24.02	24.04	24.48	25.18	23.04	25.41	**25.85**	22.48	22.53	**22.88**	21.21	21.82	22.83	22.87
House	21.66	21.70	20.05	**24.90**	21.64	23.16	22.96	20.47	20.36	18.03	**20.59**	20.28	20.58	20.38
Lena	24.69	24.76	25.21	25.55	24.04	26.19	**26.38**	22.84	22.88	22.82	21.09	22.32	**23.50**	23.48
Peppers	23.89	23.95	23.98	24.71	22.73	25.04	**25.29**	22.12	22.14	22.01	21.01	21.10	22.36	**22.48**
Kodak image1	24.40	24.46	25.49	25.02	24.19	25.59	**25.94**	22.65	22.67	23.06	21.11	22.56	23.09	**23.14**
Kodak image2	26.74	26.74	26.03	26.44	26.23	26.86	**27.36**	24.65	24.62	24.13	22.00	24.73	24.68	**24.84**
Kodak image3	26.39	26.28	25.52	26.18	25.49	26.58	**27.10**	23.98	23.85	22.92	22.07	23.69	24.15	**24.30**
Kodak image12	26.84	26.83	27.06	26.22	26.55	27.17	**27.82**	24.61	24.61	24.49	21.92	24.75	24.79	**24.99**
Average	24.36	24.40	24.39	25.16	23.77	25.38	**25.68**	22.61	22.60	22.26	21.19	22.25	22.90	**22.95**

TABLE V
RESULTS ON VIDEO DATA WITH GAUSSIAN NOISE AND RANDOM-VALUED IMPULSE NOISE

(δ, c)	Video sequence	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
$(5, 10\%)$	templete	23.24	24.12	23.88	24.09	24.26	25.05	**25.18**
	grandma	28.07	29.98	31.20	32.32	32.07	32.55	**32.74**
	akiyo	27.02	28.40	29.46	29.94	30.53	31.04	**31.28**
	bus	23.20	22.81	20.94	20.58	23.45	24.52	**24.75**
	mobile	21.07	21.46	20.39	18.01	21.48	22.73	**23.03**
	bridge-close	27.06	28.52	29.75	31.09	30.78	31.17	**31.31**
	bridge-far	30.87	31.89	33.49	35.36	34.82	35.06	**35.77**
	Average	25.79	26.74	27.01	27.34	28.19	28.87	**29.15**
$(10, 20\%)$	templete	19.71	20.67	20.85	20.66	21.09	21.44	**21.52**
	grandma	25.44	26.71	28.21	28.72	28.42	28.77	**29.27**
	akiyo	24.20	24.82	26.18	27.60	27.16	27.52	**27.99**
	bus	19.51	20.44	19.73	19.95	21.41	**21.85**	21.73
	mobile	17.88	18.52	18.64	17.39	21.15	**20.84**	20.18
	bridge-close	24.88	25.92	28.07	29.43	28.59	29.00	**29.68**
	bridge-far	28.19	28.84	31.97	32.21	31.48	31.52	**32.59**
	Average	22.83	23.70	24.80	25.13	25.61	25.84	**26.13**

including SNN, Liu, SRALT-ℓ_p, KBR, and TRPCA on the video sequences contaminated by mixed noise to demonstrate the effectiveness of the proposed model. In this experiment, (δ, c) is set to $(\delta, c) = \{(5, 10\%), (10, 20\%)\}$. Seven wildly used test videos are taken from the YUV Video Sequences to form the testing video set,[7] including templete, grandma, akiyo, bus, mobile, bridge-close, and bridge-far. The size of each frame is 144×176, and only the first 30 frames of each video are chosen for testing.

All the methods are also evaluated by the PSNR value and visual results, and the evaluation results are listed in Table V and Figs. 5 and 6. From these results, the following observations can be obtained.

1) The methods based on t-product (including TRPCA, TRWTANN, and TRWTAR-ℓ_p) obtains better results than other methods (including SNN, Liu, SRALT-ℓ_p, and KBR) in the case of video denoising. As shown in Figs. 5 and 6, the methods based on t-product retain more information of the video. This is because all the three methods based on t-product have a recovery guarantee. Also, they can find the low-rank subspace of tensor data more exactly and utilize the information within the real data more effectively than other low-rank tensor recovery methods under mixed noise.
2) The proposed methods (including TRWTANN and TRWTAR-ℓ_p) are more effective than other comparing five methods. Especially, TRWTAR-ℓ_p outperforms other comparing methods by at least 0.5 dB on average PSNR value. This indicates that the proposed methods

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

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

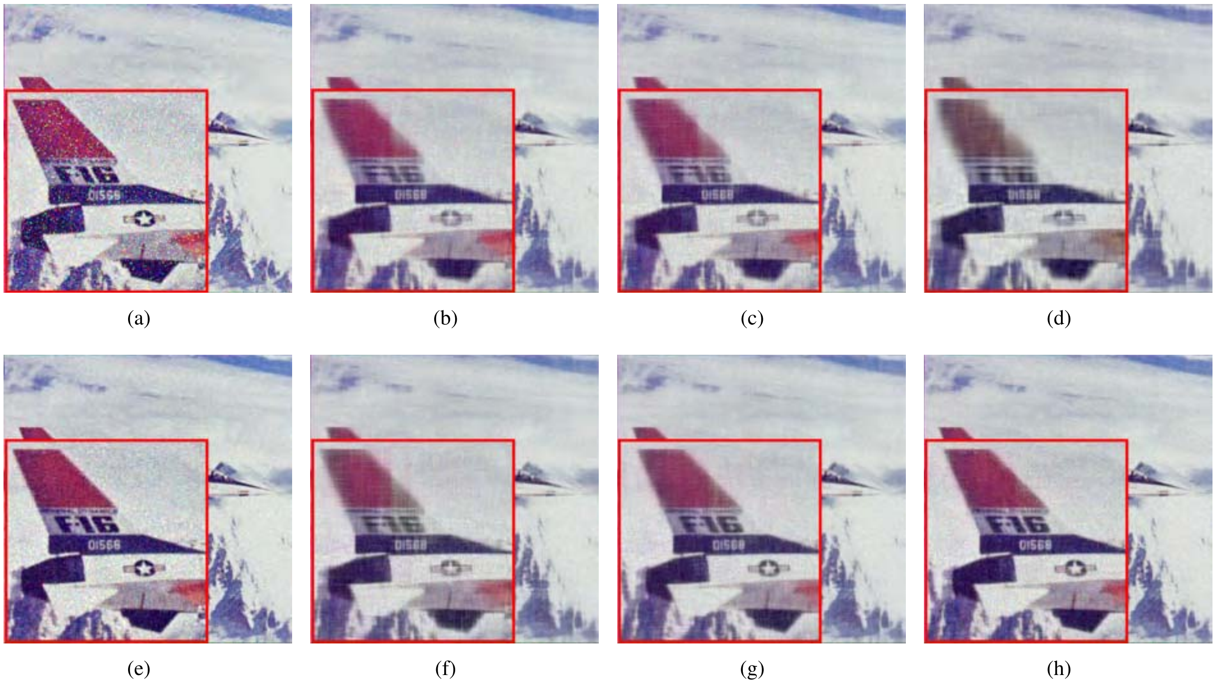

Fig. 3. Denoised results on "F16," $(\delta, c) = (15, 10\%)$. (a) Noised image. (b) SNN. (c) Liu. (d) SRALT-ℓ_p. (e) KBR. (f) TRPCA. (g) TRWTANN. (h) TRWTAR-ℓ_p.

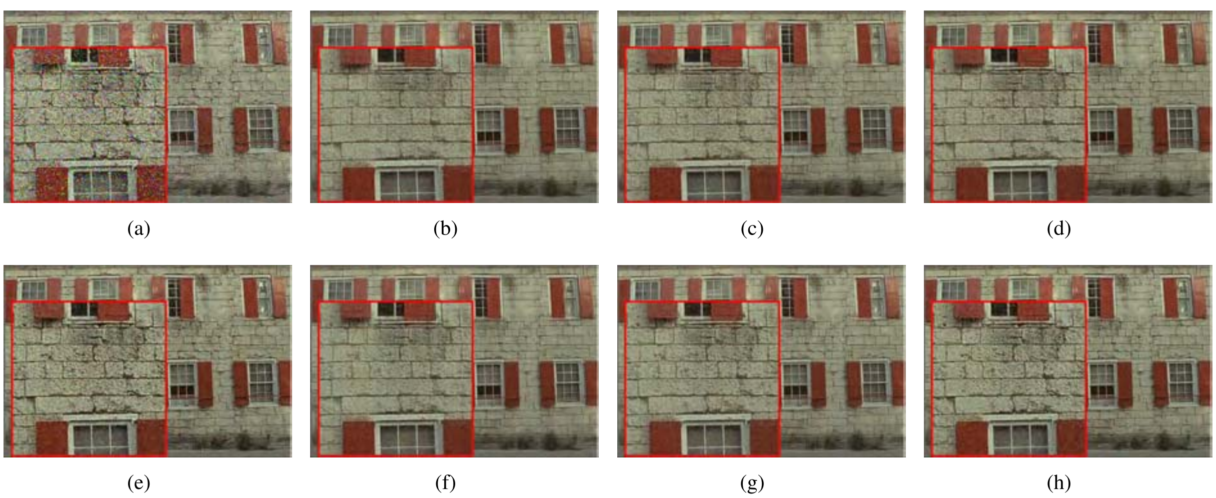

Fig. 4. Denoised results on "Kodak image1," $(\delta, c) = (15, 10\%)$. (a) Noised image. (b) SNN. (c) Liu. (d) SRALT-ℓ_p. (e) KBR. (f) TRPCA. (g) TRWTANN. (h) TRWTAR-ℓ_p.

guarantee a more accurate low-rank recovery than other comparing methods, and they are more robust against noise and outliers. 3) In most cases, the results obtained by TRWTAR-ℓ_p are better than those obtained by TRWTANN, indicating the effectiveness of the general algorithm given in this article.

D. Background Subtraction

In this part, the proposed models are applied to the background subtraction task that aims to separate the foreground objects from the background. The background of each frame of the video is static and similar, and it can be regarded as a low-rank tensor $\mathcal{L}_0$. Meanwhile, the moving foreground objects can be regarded as sparse noise $\mathcal{S}_0$, because they occupy only a fraction of pixels in the video. Therefore, all the seven methods including SNN, Liu, SRALT-ℓ_p, KBR, TRPCA, TRWTANN, and TRWTAR-ℓ_p are tested on the five video sequences[8] to deal with the case of background subtraction.

To measure the background modeling output quantitatively, $S(A, B) = ((A \bigcap B)/(A \bigcup B))$ is used to calculate the similarity between the estimated foreground regions and the ground truths. The quantitative results of different methods are listed in Table VI, and it can be seen that the proposed model achieves the best results. Also, the following observations can be made. First, TRPCA performs poorly in this experiment. This is because the exact recovery [12] and the

[8]http://perception.i2r.a-star.edu.sg/bkmodel/bkindex.html

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

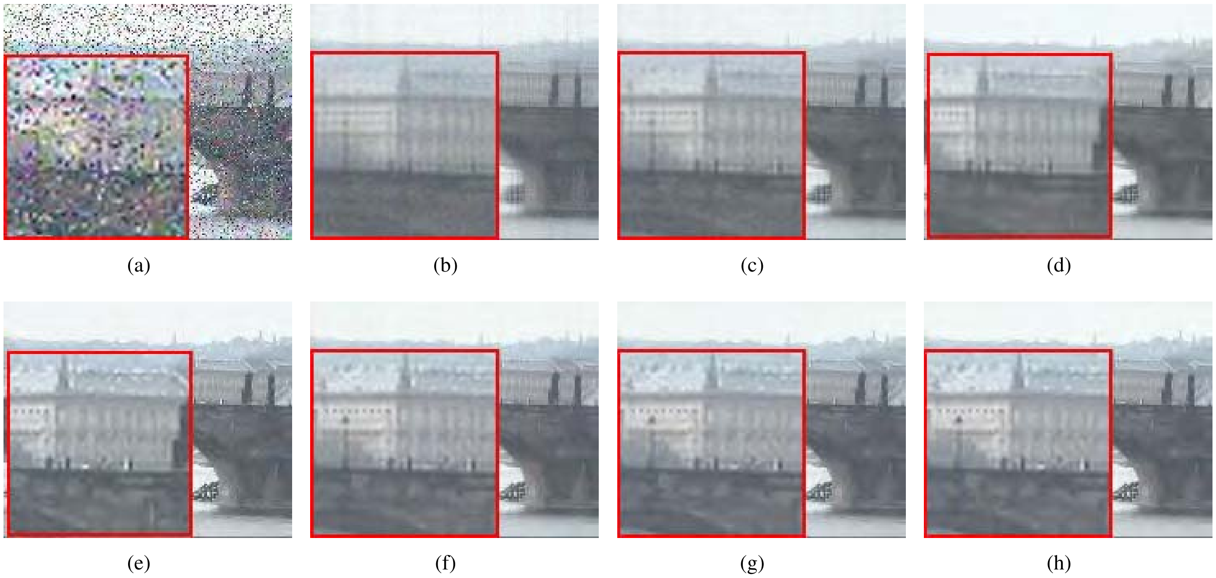

Fig. 5. Denoised results on "bridge-close," $(\delta, c) = (10, 20\%)$. (a) Noised data. (b) SNN. (c) Liu. (d) SRALT-ℓ_p. (e) KBR. (f) TRPCA. (g) TRWTANN. (h) TRWTAR-ℓ_p.

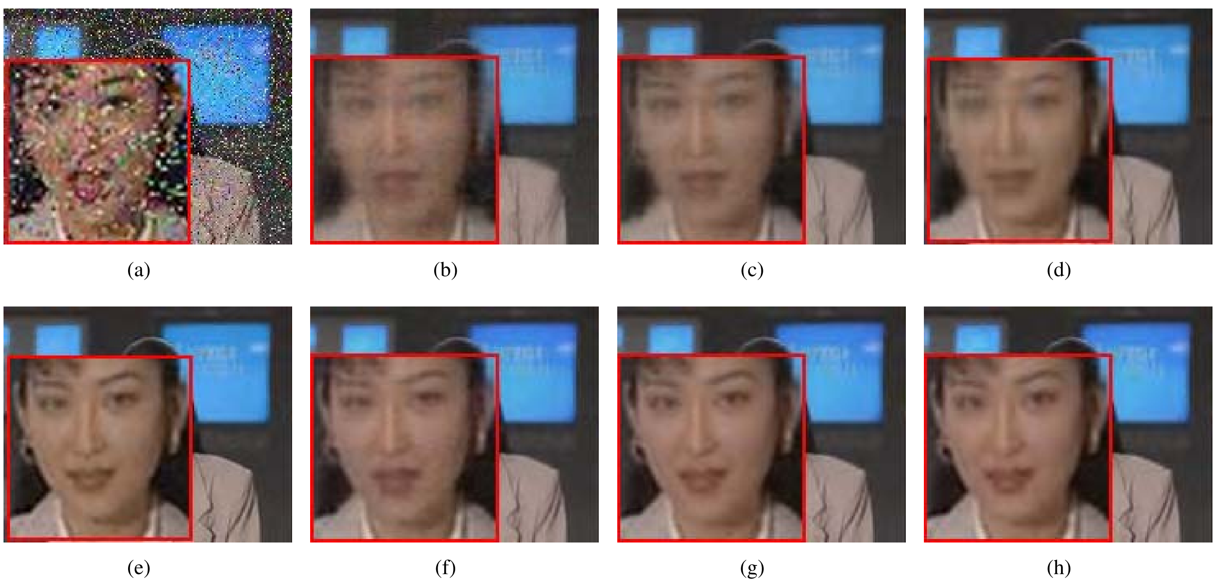

Fig. 6. Denoised results on "akiyo," $(\delta, c) = (10, 20\%)$. (a) Noised data. (b) SNN. (c) Liu. (d) SRALT-ℓ_p. (e) KBR. (f) TRPCA. (g) TRWTANN. (h) TRWTAR-ℓ_p.

TABLE VI
BACKGROUND SUBTRACTION RESULTS OF DIFFERENT METHODS

Video Clip	SNN	Liu	SRALT-ℓ_p	KBR	TRPCA	TRWTANN	TRWTAR-ℓ_p
Airport	0.2632	0.2787	0.3485	0.0859	0.1972	0.3770	**0.3837**
Hall	0.4258	0.5440	0.5408	**0.5548**	0.4412	0.5534	0.5492
Office	0.3158	0.5278	0.5081	**0.5763**	0.1874	0.5552	0.5736
Pedestrian	0.2957	**0.4882**	0.4661	0.4124	0.3177	0.4554	0.4546
Smoke	0.1138	**0.6249**	0.6063	0.5160	0.0233	0.5515	0.5881
Average	0.2829	0.4927	0.4940	0.4291	0.2334	0.4985	**0.5098**

stable recovery of TRPCA require that the support Ω of the true latent sparse tensor is uniformly distributed. However, this condition is not met in the background subtraction application because the moving foreground objects are composed of several contiguous regions. The proposed methods (TRWTANN and TRWTAR-ℓ_p) can fix this problem well. This is because they consider different transpose operators to make use of the information within the tensor data effectively, and they perform stably against the outliers. In addition, it should be noted that Liu needs some additional effort to tune the

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

weighted parameters empirically. By contrast, in our methods, all of $\alpha_k (1 \leq k \leq 3)$ are set to $(1/3)$ so that the proposed methods can be applied to real applications more easily.

VII. Conclusion and Future Work

In this work, TVTR is discussed at first. It is discovered that if different transpose operators are performed on the observation tensor, different results will be obtained by the tensor recovery algorithm with TVTR property. To solve this issue, TRWTANN is taken to study the resulting tensor by a series of transpose operators on the observation stensor, and the information within the tensor data is utilized more effectively. Besides, to balance the solvability and effectiveness of TRWTANN, the nonconvex version (10) of TRWTANN, i.e., TRWTAR-ℓ_p, is investigated. Then, the worst case error bounds of the recovered tensor are given, and a nonconvex optimization algorithm based on generalized tensor singular value thresholding (GTSVT) is designed to solve the proposed model (12) and its nonconvex version (10). The experimental results validate the effectiveness of the proposed methods.

The existing definition of tensor product-based tensor rank is limited to the order of 3. However, it is unreasonable since lots of data have more than three orders of tensor such as color video. This issue will be solved in our future work.

Appendix

A. Proof of Theorem 4

Lemma 1 [36]: Let $w_1, w_2, \ldots, w_n$ be n positive numbers such that $\sum_{k=1}^{n} w_k = 1$. Then, for any real numbers s and t such that $0 < s < t < \infty$, and for any $a_1, \ldots, a_n \geq 0$, we have

$$\left(\sum_{k=1}^{n} w_k a_k^s\right)^{\frac{1}{s}} \leq \left(\sum_{k=1}^{n} w_k a_k^t\right)^{\frac{1}{t}} \tag{24}$$

if and only if $a_1 = a_2 = \cdots = a_n$.

Theorem 4: Let $(\mathcal{L}_0, \mathcal{S}_0)$ be the pair of true low-rank and sparse tensors, and $\mathcal{L}^*$ be the solution to the optimization problem (11). If the average of the entries of the sparse component $\mathcal{S}_0$ is bounded by T, and the cardinality of the support $\mathcal{S}_0$ is bounded by m, then $\mathbf{Err}(\mathcal{L}^*) = ((\|\mathcal{L}_0 - \mathcal{L}^*\|_F)/M) \leq \sqrt[p]{((2mT^p + ((2\delta)^p/(M^{(p/2)-1})))/(M^p(1-(1/\lambda))))}$, where $\lambda > 1$, $M = \prod_{k=1}^{3} n_k$. Remark $n^{(1)} = n_2$, $n^{(2)} = n_1$, and $n^{(3)} = n_3$.

Proof: Let $(\mathcal{L}^*, \mathcal{S}^*)$ be the optimal solution of (11), $\mathcal{Z}^* = \mathcal{P} - \mathcal{S}^* - \mathcal{L}^*$, and $\mathcal{Z}_0 = \mathcal{P} - \mathcal{S}_0 - \mathcal{L}_0$. By optimality, we have

$$\|\mathcal{L}^*\|_{p,\mathrm{wa}}^p + \lambda\|\mathcal{P} - \mathcal{Z}^* - \mathcal{L}^*\|_{p,p}^p \leq \|\mathcal{L}_0\|_{p,\mathrm{wa}}^p + \lambda\|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p. \tag{25}$$

Next, recall that a function $f(\cdot)$ is sub-additive if $f(x+y) \leq f(x) + f(y)$. According to the result in [37], a concave function $f : [0,\infty) \to [0,\infty)$ with $f(0) \geq 0$ is sub-additive. Thus, for $0 < p < 1$, $f(x) = |x|^p$ is concave (x is a scalar here), $|x|^p$ is a sub-additive function. Since the sum of sub-additive functions is sub-additive, $f(x) = \|x\|_p^p, x \in \mathbb{R}^n$ is also sub-additive, thereby implying $\|x\|_p^p - \|y\|_p^p \leq \|x - y\|_p^p$. Consequently, (25) implies that

$$\begin{aligned}
&\|\mathcal{P} - \mathcal{Z}^* - \mathcal{L}^*\|_{p,p}^p \\
&\quad\leq \frac{1}{\lambda}\left(\|\mathcal{L}_0\|_{p,\mathrm{wa}}^p - \|\mathcal{L}^*\|_{p,\mathrm{wa}}^p\right) + \|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p \\
&\quad\leq \frac{1}{\lambda}\|\mathcal{L}_0 - \mathcal{L}^*\|_{p,\mathrm{wa}}^p + \|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p
\end{aligned} \tag{26}$$

where the last inequality is derived from the linearity property in the definition of $\|\cdot\|_{p,\mathrm{wa}}^p$ on tensors. Based on this inequality, $\|\mathcal{L}^* - \mathcal{L}_0\|_{p,p}^p$ can be bounded as follows:

$$\begin{aligned}
&\|\mathcal{L}_0 - \mathcal{L}^*\|_{p,p}^p \\
&\leq \|\mathcal{P} - \mathcal{Z}^* - \mathcal{L}^*\|_{p,p}^p + \|\mathcal{P} - \mathcal{Z}^* - \mathcal{L}_0\|_{p,p}^p \\
&\leq \|\mathcal{P} - \mathcal{Z}^* - \mathcal{L}^*\|_{p,p}^p + \|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p + \|\mathcal{Z}^* - \mathcal{Z}_0\|_{p,p}^p \\
&\leq \frac{1}{\lambda}\|\mathcal{L}_0 - \mathcal{L}^*\|_{p,\mathrm{wa}}^p + 2\|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p + \|\mathcal{Z}^* - \mathcal{Z}_0\|_{p,p}^p \\
&= \frac{1}{\lambda}\sum_{k=1}^{3}\frac{\alpha_k}{n^{(k)}}\left(\sum_{i=1}^{r_k}\left(\sigma_i^{(k)}\right)^p\right) + 2\|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p \\
&\quad + \|\mathcal{Z}^* - \mathcal{Z}_0\|_{p,p}^p
\end{aligned} \tag{27}$$

where the third inequality is derived from substituting the inequality (26) into the current inequality; r_k is the rank of the matrix $\mathrm{bcirc}((\mathcal{L}_0 - \mathcal{L}^*)^{T_k})$, and $\sigma_1^{(k)}, \sigma_2^{(k)}, \ldots, \sigma_{r_k}^{(k)}$ are the r_k singular values of the matrix $\mathrm{bcirc}((\mathcal{L}_0 - \mathcal{L}^*)^{T_k})$.

Since $\|Z_0\|_F \leq \delta$ and $\|Z^*\|_F \leq \delta$, $\|Z_0 - Z^*\|_F \leq 2\delta$. According to Lemma 1 and setting $w_j = (1/M), \forall j = 1, \ldots, M$, we have

$$\|\mathcal{Z}^* - \mathcal{Z}_0\|_{p,p}^p \leq M\left(\frac{\|Z^* - Z_0\|_F}{\sqrt{M}}\right)^p \leq \frac{(2\delta)^p}{M^{\frac{p}{2}-1}}.$$

By Lemma 1, and setting $w_j = (1/r_k), \forall j = 1, \ldots, r_k$, we have

$$\begin{aligned}
&\frac{\left(\sigma_1^{(k)}\right)^p + \left(\sigma_2^{(k)}\right)^p + \cdots + \left(\sigma_{r_k}^{(k)}\right)^p}{r_k} \\
&\qquad\leq \left(\sqrt{\frac{\left(\sigma_1^{(k)}\right)^2 + \left(\sigma_2^{(k)}\right)^2 + \cdots + \left(\sigma_{r_k}^{(k)}\right)^2}{r_k}}\right)^p
\end{aligned} \tag{28}$$

thereby leading to

$$\begin{aligned}
\sum_{i=1}^{r_k}\left(\sigma_i^{(k)}\right)^p &\leq r_k^{1-\frac{p}{2}}\left(\sum_{i=1}^{r_k}\left(\sigma_i^{(k)}\right)^2\right)^{\frac{p}{2}} \\
&= r_k^{1-\frac{p}{2}}\|\mathrm{bcirc}((\mathcal{L}_0 - \mathcal{L}^*)^{T_k})\|_F^p \\
&= r_k^{1-\frac{p}{2}}(n^{(k)})^{\frac{p}{2}}\|\mathcal{L}_0 - \mathcal{L}^*\|_F^p.
\end{aligned} \tag{29}$$

Applying this inequality to the final line in (27) results in

$$\begin{aligned}
\|\mathcal{L}^* - \mathcal{L}_0\|_{p,p}^p &\leq \frac{1}{\lambda}\sum_{k=1}^{3}\frac{\alpha_k}{(n^{(k)})^{1-\frac{p}{2}}}r_k^{1-\frac{p}{2}}\|\mathcal{L}_0 - \mathcal{L}^*\|_F^p \\
&\quad + 2mT^p + \frac{(2\delta)^p}{M^{\frac{p}{2}-1}}.
\end{aligned} \tag{30}$$

Since $\|\mathcal{P} - \mathcal{Z}_0 - \mathcal{L}_0\|_{p,p}^p = \|\mathcal{S}_0\|_{p,p}^p \leq mT^p$, according to the generalized power-mean inequality in Lemma 1 (by setting

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$s = p, t = 1$), we have

$$\left(\frac{\|\mathcal{S}_0\|_{p,p}^p}{m}\right)^{\frac{1}{p}} \le \frac{\|\mathcal{S}_0\|_{1,1}^1}{m} \le T. \tag{31}$$

Next, we show that $\|\mathcal{L}_0 - \mathcal{L}^*\|_F^p \le \|\mathcal{L}^* - \mathcal{L}_0\|_{p,p}^p$. Denoting $\mathcal{L} = \mathcal{L}_0 - \mathcal{L}^*$ and based on the fact that $\|\mathcal{L}_0 - \mathcal{L}^*\|_F \le \|\mathcal{L}_0 - \mathcal{L}^*\|_1$, we have

$$\begin{aligned}\|\mathcal{L}\|_F &= \sqrt{\sum_{i_1,i_2,i_3} \mathcal{L}_{i_1i_2i_3}^2} \le \sum_{i_1,i_2,i_3} |\mathcal{L}_{i_1i_2i_3}| \\ &= \left(\left(\sum_{i_1,i_2,i_3} |\mathcal{L}_{i_1i_2i_3}|\right)^p\right)^{\frac{1}{p}} \le \left(\sum_{i_1,i_2,i_3} |\mathcal{L}_{i_1i_2i_3}|^p\right)^{\frac{1}{p}} \\ &= \|\mathcal{L}\|_{p,p}\end{aligned} \tag{32}$$

where the second inequality is derived from the fact that $f(x) = x^p (0 < p < 1)$ is a sub-additive function. Raising both sides to the power of p yields $\|\mathcal{L}_0 - \mathcal{L}^*\|_F^p \le \|\mathcal{L}^* - \mathcal{L}_0\|_{p,p}^p$. Combining this inequality with (30), we have

$$\begin{aligned}\|\mathcal{L}_0 - \mathcal{L}^*\|_F^p \le \frac{1}{\lambda}\sum_{k=1}^{3} \frac{\alpha_k}{(n^{(k)})^{1-\frac{p}{2}}} r_k^{1-\frac{p}{2}} \|\mathcal{L}_0 - \mathcal{L}^*\|_F^p \\ + 2mT^p + \frac{(2\delta)^p}{M^{\frac{p}{2}-1}}.\end{aligned} \tag{33}$$

Rearranging the terms, we have $\|\mathcal{L}_0 - \mathcal{L}^*\|_F^p \le ((2mT^p + ((2\delta)^p/(M^{(p/2)-1})))/(1 - (1/\lambda)\sum_{k=1}^3 (\alpha_k/((n^{(k)})^{1-(p/2)}))$ $r_k^{1-(p/2)})) \le ((2mT^p + ((2\delta)^p/(M^{(p/2)-1})))/(1 - (1/\lambda)\sum_{k=1}^3$ $(\alpha_k/((n^{(k)})^{1-(p/2)}))(n^{(k)})^{1-(p/2)}))$ (since $r_k \le n^{(k)}$ and $1 - (p/2) > 0$), and therefore

$$\|\mathcal{L}_0 - \mathcal{L}^*\|_F \le \sqrt[p]{\frac{2mT^p + \frac{(2\delta)^p}{M^{\frac{p}{2}-1}}}{1 - \frac{1}{\lambda}}} \tag{34}$$

provided that $\lambda > 1$. ∎

B. Proof of Theorem 5

Lemma 2 [38]: Given any real vector $\upsilon \in \mathbb{R}^n$, the associated $\bar{\upsilon} = F_n\upsilon \in \mathbb{C}^n$ satisfies

$$\bar{\upsilon}_1 \in \mathbb{R} \quad \text{and} \quad \mathrm{Conj}(\bar{\upsilon}_i) = \bar{\upsilon}_{n-i+2}, \quad i = 2, \ldots, \left\lfloor \frac{n+1}{2} \right\rfloor. \tag{35}$$

Conversely, for any given complex $\bar{\upsilon} \in \mathbb{C}^n$ that satisfies (35), there exists a real block circulant matrix $\mathrm{circ}(\upsilon)$ such that $F_n \mathrm{circ}(\upsilon) F_n^{-1} = \mathrm{Diag}(\bar{\upsilon})$ holds.

Lemma 3 [39], [40]: For any matrices $A, B \in \mathbb{C}^{m\times n} (m \le n)$, $\mathrm{Re}(\mathrm{tr}(AB^*)) \le \Sigma_{i=1}^m \sigma_i(A)\sigma_i(B)$, where $\sigma_1(A) \ge \sigma_2(A) \ge \cdots \ge 0$ and $\sigma_1(B) \ge \sigma_2(B) \ge \cdots \ge 0$ are the singular values of A and B, respectively. The equality holds if and only if there exist unitary matrices U and V such that $A = U\mathrm{diag}(\sigma(A))V^*$ and $B = U\mathrm{diag}(\sigma(B))V^*$ are the SVDs of A and B.

Lemma 4 [41]: For any lower bounded function g, $p_1^* \ge p_2^*$ if $x_1 \ge x_2$ for any $p_i^* \in \mathrm{prox}_{\lambda,g}(x_i), i = 1, 2$, where $\mathrm{prox}_{\lambda,g}(y) = \arg\min_{x\in\mathbb{R}} (1/2)(y - x)^2 + \lambda g(|x|)$.

Lemma 5: For any two tensors $\bar{\mathcal{A}}, \bar{\mathcal{B}} \in \{\bar{\mathcal{M}} \in \mathbb{C}^{n_1\times n_2\times n_3} | \bar{M}^{(1)} \in \mathbb{R}^{n_1\times n_2}; \mathrm{Conj}(\bar{M}^{(i)}) = \bar{M}^{(n_3-i+2)}, i = 2, \ldots, \lfloor((n_3 + 1)/2)\rfloor.\}$, then $\mathrm{Im}(\langle \mathrm{bdiag}(\bar{\mathcal{A}}), \mathrm{bdiag}(\bar{\mathcal{B}})\rangle) = 0$.

Proof: According to Lemma 2, there exist two real block circulant matrices $\mathrm{bcirc}(\mathcal{A})$ and $\mathrm{bcirc}(\mathcal{B})$ such that $\mathrm{bdiag}(\bar{\mathcal{A}}) = (F_{n_3} \otimes I_{n_1}) \cdot \mathrm{bcirc}(\mathcal{A}) \cdot (F_{n_3}^{-1} \otimes I_{n_2})$ and $\mathrm{bdiag}(\bar{\mathcal{B}}) = (F_{n_3} \otimes I_{n_1}) \cdot \mathrm{bcirc}(\mathcal{B}) \cdot (F_{n_3}^{-1} \otimes I_{n_2})$. Thus $\mathrm{Im}(\langle \mathrm{bdiag}(\bar{\mathcal{A}}), \mathrm{bdiag}(\bar{\mathcal{B}})\rangle) = \mathrm{Im}(\langle \mathrm{bcirc}(\mathcal{A}), \mathrm{bcirc}(\mathcal{B})\rangle) = 0$. ∎

Lemma 6: Let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be a function such that for $x \in \mathbb{R}$, $g(|x|) \ge 0$, and $g(0) = 0$. Let $Y = U\mathrm{diag}(\sigma(Y))V^*$ be the SVD of $Y \in \mathbb{C}^{m\times n}$, then

$$\begin{aligned}\arg \min_{\mathrm{Im}(\langle X,Y\rangle)=0, X\in\mathbb{C}^{m\times n}} \frac{1}{2}\|Y - X\|_F^2 + \lambda \sum_{i=1}^{\min(m,n)} g(\sigma_i(X)) \\ = \{U\Sigma V^* | \Sigma \in \mathrm{diag}(\mathcal{T}_g(\sigma(Y), \lambda))\}\end{aligned} \tag{36}$$

where, for $\mathrm{y} \in \mathbb{R}^h$

$$\mathcal{T}_g(\mathrm{y}, \lambda) = \arg\min_{\mathrm{x}\in\mathbb{R}^h} \frac{1}{2}\|\mathrm{y} - \mathrm{x}\|_F^2 + \lambda \sum_{i=1}^{h} g(|\mathrm{x}_i|).$$

Proof: For the convenience of discussion, remark that

$$\mathbf{A} = \arg \min_{\mathrm{Im}(\langle X,Y\rangle)=0, X\in\mathbb{C}^{m\times n}} \frac{1}{2}\|Y - X\|_F^2 + \lambda \sum_{i=1}^{\min(m,n)} g(\sigma_i(X))$$

and $\mathbf{B} = \{U\Sigma V^* | \Sigma \in \mathrm{diag}(\mathcal{T}_g(\sigma(Y), \lambda))\}$.

For any two complex matrices X and Y that satisfy $\mathrm{Im}(\langle X, Y\rangle) = 0$, we have

$$\begin{aligned}\|Y - X\|_F^2 &= \|X\|_F^2 + \|Y\|_F^2 - 2\mathrm{tr}(XY^*) \\ &= \|X\|_F^2 + \|Y\|_F^2 - 2\mathrm{Re}(\mathrm{tr}(XY^*)) \\ &\ge \|\sigma(X)\|_2^2 + \|\sigma(Y)\|_2^2 - 2\Sigma_{i=1}^{\min(m,n)} \sigma_i(X)\sigma_i(Y) \\ &= \|\sigma(X) - \sigma(Y)\|_2^2.\end{aligned} \tag{37}$$

According to Lemma 3, the equality (37) holds if U and V are left singular value vector matrix and right singular value vector matrix of X, respectively. In this case, the optimal problem (36) reduces to

$$\arg \min_{\mathrm{X}:\mathrm{X}_1\ge\cdots\ge\mathrm{X}_{\min(m,n)}\ge 0} \sum_{i=1}^{\min(m,n)} \left(\lambda g(|\mathrm{x}_i|) + \frac{1}{2}(\mathrm{x}_i - \sigma_i(Y))^2\right). \tag{38}$$

Since $\sigma_1(Y) \ge \sigma_2(Y) \ge \cdots \ge \sigma_{\min(m,n)}(Y)$, according to Lemma 4, there exits $\hat{\mathrm{x}} \in \mathcal{T}_g(\sigma(Y), \lambda)$ such that $\hat{\mathrm{x}}_1 \ge \hat{\mathrm{x}}_2 \cdots \ge \hat{\mathrm{x}}_{\min(m,n)}$. Such a choice of $U\mathrm{diag}(\hat{\mathrm{x}})V^*$ is an optimal solution of (36), so $\mathbf{A} \supseteq \mathbf{B}$ holds. $\mathbf{A} \subseteq \mathbf{B}$ is proven as follows.

If there $\hat{X}$ belongs to $\mathbf{A}$ but not belongs to $\mathbf{B}$, then according to Lemma 3, we have $(1/2)\|Y-\hat{X}\|_F^2+\lambda\sum_{i=1}^{\min(m,n)} g(\sigma_i(\hat{X})) > (1/2)\|\sigma(Y) - \sigma(\hat{X})\|_2^2 + \lambda\sum_{i=1}^{\min(m,n)} g(\sigma_i(\hat{X})) = (1/2)\|Y - U\mathrm{diag}(\sigma(\hat{X}))V^*\|_F^2 + \lambda\sum_{i=1}^{\min(m,n)} g(\sigma_i(\hat{X}))$. Note that $\mathrm{Im}(\langle U\mathrm{diag}(\sigma(\hat{X}))V^*, Y\rangle) = 0$, so the following expression is a contradiction to $\hat{X} \in \mathbf{A}$:

$$\begin{aligned}&\frac{1}{2}\|Y - \hat{X}\|_F^2 + \lambda \sum_{i=1}^{\min(m,n)} g(\sigma_i(\hat{X})) \\ &\quad > \min_{\mathrm{Im}(\langle X,Y\rangle)=0, X\in\mathbb{C}^{m\times n}} \frac{1}{2}\|Y - X\|_F^2 + \lambda \sum_{i=1}^{\min(m,n)} g(\sigma_i(X)).\end{aligned}$$

Therefore, $\mathbf{A} = \mathbf{B}$. ∎

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

The following definitions are given before the proof of Theorem 5 is presented.

1) $\mathbf{M} = \{\text{bdiag}(\bar{\mathcal{M}})|\bar{\mathcal{M}} \in \mathbb{C}^{n_1\times n_2\times n_3}, \bar{M}^{(1)} \in \mathbb{R}^{n_1\times n_2}.\text{Conj}(\bar{M}^{(i)}) = \bar{M}^{(n_3-i+2)}, i = 2,\ldots,\lfloor((n_3+1)/2)\rfloor\}$.
2) Let $\mathcal{S}_{\lambda,g}, \bar{\mathcal{S}}_{\lambda,g} \in \mathbb{R}^{r\times r\times n_3}$, $\bar{S}_{\lambda,g}^{(i)} = \mathcal{T}_g(\bar{S}^{(i)},\lambda)$ and $\bar{S}_{\lambda,g} = \text{bdiag}(\bar{\mathcal{S}}_{\lambda,g}) = (F_{n_3}\otimes I_r)\cdot\text{bcirc}(\mathcal{S}_{\lambda,g})\cdot(F_{n_3}^{-1}\otimes I_r)$.

Theorem 5: For any $\lambda > 0$ and $\mathcal{Y} \in \mathbb{R}^{n_1\times n_2\times n_3}$, if g is increasing on $[0,+\infty)$, then the tensor singular value thresholding operator obeys

$$\mathcal{D}_g(\mathcal{Y},\lambda) \in \arg\min_{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}} \lambda\|\mathcal{X}\|_{*,a}^{g} + \frac{1}{2}\|\mathcal{Y}-\mathcal{X}\|_F^2. \quad (39)$$

Proof: It is shown first that $\mathcal{D}_g(\mathcal{Y},\lambda)$ is a real tensor. Assume $\mathcal{U} *_t \mathcal{S} *_t \mathcal{V}^*$ is the t-SVD of $\mathcal{Y}$. Then, according to Theorem 1, $\mathcal{U}$ and $\mathcal{V}$ are real. Since $\bar{\mathcal{S}}$ is real, and $\bar{S} \in \mathbf{M}$, we have $\bar{S}_{\lambda,g} \in \mathbf{M}$. According to Lemma 2, there exists a real black circulant matrix $\text{bcirc}(\mathcal{S}_{\lambda,g})$ such that $\bar{S}_{\lambda,g} = (F_{n_3}\otimes I_r)\cdot \text{bcirc}(\mathcal{S}_{\lambda,g})\cdot(F_{n_3}^{-1}\otimes I_r)$. Therefore, $\mathcal{D}_g(\mathcal{Y},\lambda) = \mathcal{U} *_t \mathcal{S}_{\lambda,g} *_t \mathcal{V}^*$ is real.

According to the definition of $\mathcal{D}_g(\mathcal{Y},\lambda)$ and Lemma 6, $(F_{n_3}\otimes I_{n_1})\cdot\text{bcirc}(\mathcal{D}_g(\mathcal{Y},\lambda))\cdot(F_{n_3}^{-1}\otimes I_{n_2}) \in \arg\min_{\text{Im}(\langle X,\bar{Y}\rangle)=0}(1/2)\|\bar{Y}-X\|_F^2 + \lambda\sum_{i=1}^{\min(n_1,n_2)n_3} g(\sigma_i(X))$. On the other hand, since $(F_{n_3}\otimes I_{n_1})\cdot\text{bcirc}(\mathcal{D}_g(\mathcal{Y},\lambda))\cdot(F_{n_3}^{-1}\otimes I_{n_2}) \in \mathbf{M}$, $(F_{n_3}\otimes I_{n_1})\cdot\text{bcirc}(\mathcal{D}_g(\mathcal{Y},\lambda))\cdot(F_{n_3}^{-1}\otimes I_{n_2}) \in \arg\min_{\text{Im}(\langle X,\bar{Y}\rangle)=0, X\in\mathbf{M}}(1/2)\|\bar{Y}-X\|_F^2 + \lambda\sum_{i=1}^{\min(n_1,n_2)n_3} g(\sigma_i(X)) = \arg\min_{X\in\mathbf{M}}(1/2)\|\bar{Y}-X\|_F^2 + \lambda\sum_{i=1}^{\min(n_1,n_2)n_3} g(\sigma_i(X))$, where the second equation holds according to Lemma 5 (Note that $\bar{Y} \in \mathbf{M}$). Therefore, $\mathcal{D}_g(\mathcal{Y},\lambda) \in \arg\min_{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}} \lambda\|\mathcal{X}\|_{*,a}^g + (1/2)\|\mathcal{Y}-\mathcal{X}\|_F^2$. ■

References

[1] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2009.

[2] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Apr. 2009.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[4] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions," *IFAC Proc. Volumes*, vol. 42, no. 10, pp. 1493–1498, 2009.

[5] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 2496–2504.

[6] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.

[7] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936.

[8] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[9] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 1518–1522.

[10] 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. 2012.

[11] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, 2011, Art. no. 025010.

[12] 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.*, vol. 42, no. 4, pp. 925–938, Jan. 2020.

[13] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 238–255, Jan. 2021.

[14] Y.-B. Zheng, T.-Z. Huang, T.-Y. Ji, X.-L. Zhao, T.-X. Jiang, and T.-H. Ma, "Low-rank tensor completion via smooth matrix factorization," *Appl. Math. Model.*, vol. 70, pp. 677–695, Jun. 2019.

[15] F. Zhang, J. Wang, W. Wang, and C. Xu, "Low-tubal-rank plus sparse tensor recovery with prior subspace information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3492–3507, Oct. 2021.

[16] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, and T.-Z. Huang, "Low-rank tensor train for tensor robust principal component analysis," *Appl. Math. Comput.*, vol. 367, Feb. 2020, Art. no. 124783.

[17] C. Cai, G. Li, H. Vincent Poor, and Y. Chen, "Nonconvex low-rank tensor completion from noisy data," 2019, *arXiv:1911.04436*.

[18] C. Lu, X. Peng, and Y. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 5996–6004.

[19] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[20] R. Tomioka, K. Hayashi, and H. Kashima, "Estimation of low-rank tensors via convex optimization," 2010, *arXiv:1010.0789*.

[21] M. Signoretto, Q. T. Dinh, L. De Lathauwer, and J. A. K. Suykens, "Learning with tensors: A framework based on convex optimization and spectral regularization," *Mach. Learn.*, vol. 94, no. 3, pp. 303–351, 2014.

[22] C. Mu, B. Huang, J. Wright, and D. Goldfarb, "Square deal: Lower bounds and improved relaxations for tensor recovery," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 73–81.

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

[24] N. Hao, L. Horesh, and M. Kilmer, "Nonnegative tensor decomposition," *Compressed Sens. Sparse Filtering*, vol. 34, no. 1, pp. 148–172, 2014.

[25] X. Zhang, Z. Zhou, D. Wang, and Y. Ma, "Hybrid singular value thresholding for tensor completion," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 1362–1368.

[26] Y. Liu and F. Shang, "An efficient matrix factorization method for tensor completion," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 307–310, Apr. 2013.

[27] L. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.

[28] D. Malioutov and A. Aravkin, "Iterative log thresholding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 7198–7202.

[29] J. H. Friedman, "Fast sparse regression and classification," *Int. J. Forecasting*, vol. 28, no. 3, pp. 722–738, Jul./Sep. 2012.

[30] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, Jan. 2009.

[31] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *J. Mach. Learn. Res.*, vol. 11, pp. 1081–1107, Mar. 2010.

[32] C. Gao, N. Wang, Q. Yu, and Z. Zhang, "A feasible nonconvex relaxation approach to feature selection," in *Proc. AAAI*, 2014, pp. 356–361.

[33] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017, doi: 10.1109/TSP.2017.2690524.

[34] D. P. Bertsekas, *Nonlinear Programming*. Beijing, China: Tsinghua University Press, 2018.

[35] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.

[36] P. S. Bullen, D. S. Mitrinovic, and M. Vasic, *Means and Their Inequalities*, vol. 31. Dordrecht, The Netherlands: Springer, 2013.

[37] E. Schechter, *Handbook of Analysis and Its Foundations*. New York, NY, USA: Academic, 1996.

[38] O. Rojo and H. Rojo, "Some results on symmetric circulant matrices and on symmetric centrosymmetric matrices," *Linear Algebra Appl.*, vol. 392, pp. 211–233, Nov. 2004.

[39] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, vol. 143. New York, NY, USA: Springer, 1979.

[40] N. Komaroff, "Enhancements to the von Neumann trace inequality," *Linear Algebra Appl.*, vol. 428, no. 4, pp. 738–741, Feb. 2008.

[41] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," 2014, *arXiv:1412.2231*.

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Xiaoqin Zhang (Senior Member, IEEE) received the B.E. degree in electronic information science and technology from Central South University, Changsha, China, in 2005, and the Ph.D. degree in pattern recognition and intelligent system from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently a Professor with Wenzhou University, Wenzhou, China. He has authored or coauthored over 80 papers in international and national journals and international conferences. His research interests include pattern recognition, computer vision, and machine learning.

Jingjing Zheng received the B.A. degree in art and design from the Wuchang Institute of Technology, Wuhan, China, in 2015, and the M.S. degree in applied mathematics from the College of Mathematics and Physics, Wenzhou University, Wenzhou, China, in 2020. She is currently pursuing the Ph.D. degree with the Department of Computer Science, Memorial University, St. John's, NL, Canada, and co-supervised by Prof. X. Zhang and Prof. X. Jiang.

Her current research interests include pattern recognition and machine learning.

Li Zhao received the B.E. degree in automation and the M.Eng. degree in control theory and control engineering from Central South University, Changsha, China, in 2005 and 2008, respectively.

She is currently an Associate Professor with Wenzhou University, Wenzhou, China. Her research interests include pattern recognition, computer vision, and machine learning.

Zhengyuan Zhou received the B.E. degree in electrical engineering and computer sciences and the B.A. degree in mathematics from the University of California at Berkeley, Berkeley, CA, USA, in 2012, and the Ph.D. degree from the Information System Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA, USA, in summer 2019.

He has substantial industry experience at Google, Microsoft Research, and Oracle. He is currently an Assistant Professor with the NYU Stern School of Business, New York, NY, USA. His research interests include learning, optimization, game theory, and stochastic systems.

Zhouchen Lin (Fellow, IEEE) received the Ph.D. degree from Peking University, Beijing, China, in 2000.

He is currently a Professor with the Key Laboratory of Machine Perception, School of Artificial Intelligence, Peking University. His research interests include machine learning and numerical optimization.

Dr. Lin is also a fellow of the International Association of Pattern Recognition (IAPR) and China Society of Image and Graphics (CSIG). He has been an Area Chair of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), Annual Conference on Neural Information Processing Systems (NIPS/NeurIPS), AAAI Conference on Artificial Intelligence (AAAI), International Joint Conference on Artificial Intelligence (IJCAI), International Conference on Learning Representations (ICLR), and International Conference on Machine Learning (ICML) many times. He was the Program Co-Chair of International Conference on Pattern Recognition (ICPR) 2022 and the Senior Area Chair of ICML 2022 and NeurIPS 2022. He is also an Associate Editor of the *International Journal of Computer Vision* and *Optimization Methods and Software*.

Authorized licensed use limited to: Peking University. Downloaded on June 24,2022 at 01:23:44 UTC from IEEE Xplore. Restrictions apply.