Information Sciences 532 (2020) 170–189

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Tensor N-tubal rank and its convex relaxation for low-rank tensor recovery

Yu-Bang Zheng [a], Ting-Zhu Huang [a,*], Xi-Le Zhao [a,*], Tai-Xiang Jiang [b], Teng-Yu Ji [c], Tian-Hui Ma [d]

[a] School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, PR China
[b] School of Economic Information Engineering/FinTech Innovation Center, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, PR China
[c] School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, PR China
[d] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China

ARTICLE INFO

Article history:
Received 21 November 2018
Received in revised form 21 April 2020
Accepted 1 May 2020
Available online 5 May 2020

Keywords:
Low-rank tensor recovery (LRTR)
Mode-k_1k_2 tensor unfolding
Tensor N-tubal rank
Weighted sum of tensor nuclear norm (WSTNN)
Alternating direction method of multipliers (ADMM)

ABSTRACT

The recent popular tensor tubal rank, defined based on tensor singular value decomposition (t-SVD), yields promising results. However, its framework is applicable only to three-way tensors and lacks the flexibility necessary tohandle different correlations along different modes. To tackle these two issues, we define a new tensor unfolding operator, named mode-k_1k_2 tensor unfolding, as the process of lexicographically stacking all mode-k_1k_2 slices of an N-way tensor into a three-way tensor, which is a three-way extension of the well-known mode-k tensor matricization. On this basis, we define a novel tensor rank, named the tensor N-tubal rank, as a vector consisting of the tubal ranks of all mode-k_1k_2 unfolding tensors, to depict the correlations along different modes. To efficiently minimize the proposed N-tubal rank, we establish its convex relaxation: the weighted sum of the tensor nuclear norm (WSTNN). Then, we apply the WSTNN to low-rank tensor completion (LRTC) and tensor robust principal component analysis (TRPCA). The corresponding WSTNN-based LRTC and TRPCA models are proposed, and two efficient alternating direction method of multipliers (ADMM)-based algorithms are developed to solve the proposed models. Numerical experiments demonstrate that the proposed models significantly outperform the compared ones.

1. Introduction

As a multidimensional array, the tensor [20] plays an increasingly significant role in many applications, such as color image/video processing [13,26,32,45], hyperspectral/multispectral image (HSI/MSI) processing [7,22,47,38], background subtraction [18,3], video rain streak removal [34,21], and magnetic resonance imaging (MRI) data recovery [15,17,37,6]. Many of these applications can be formulated as a class of tensor recovery problems, i.e., recovering an underlying tensor from its corrupted observation. Particularly, as two typical examples, tensor completion aims to complete missing elements, and tensor robust principal component analysis (TRPCA) aims to remove sparse outliers. The key to tensor recovery is to explore the redundancy prior of the underlying tensor, which is usually formulated as low-rankness. Thus, low-rank modeling has been widely studied and has achieved great success in the tensor recovery task.

* Corresponding authors.
E-mail address: tingzhuhuang@126.com (T.-Z. Huang).

https://doi.org/10.1016/j.ins.2020.05.005
0020-0255/

The traditional matrix recovery is a two-way tensor recovery problem. Since the matrix rank, measured by the number of non-zero singular values, is powerful enough to capture the global information of a matrix, most matrix recovery methods aim to minimize the matrix rank [2,1,30,5]. However, directly minimizing the matrix rank is NP-hard [11]. To tackle this issue, the nuclear norm ($\|\cdot\|_*$), i.e., the sum of all non-zero singular values, has been proposed to approximate the matrix rank, leading to great successes [2,1].

Tensor recovery can be viewed as an extension of matrix recovery. Inspired by the success of matrix rank minimization, it seems natural to recover the underlying tensor by minimizing the tensor rank. Mathematically, a general low-rank tensor recovery (LRTR) model can be written as

$$\min_{\mathcal{X}} \quad \text{rank}(\mathcal{X}) + \lambda L(\mathcal{X}, \mathcal{F}), \tag{1}$$

where $\mathcal{X}$ is the underlying tensor, $\mathcal{F}$ is the observed tensor, and $L(\mathcal{X}, \mathcal{F})$ is the loss function between $\mathcal{X}$ and $\mathcal{F}$, e.g., $\mathcal{X}_\Omega = \mathcal{F}_\Omega$ for low-rank tensor completion (LRTC) and $\|\mathcal{F} - \mathcal{X}\|_1$ for TRPCA. A conclusive issue of LRTR is the definition of the tensor rank. However, unlike the matrix rank, the definition of the tensor rank is not unique. Many research efforts have been devoted to defining the tensor rank, and most of them are defined based on the corresponding tensor decomposition, such as the CANDECOMP/PARAFAC (CP) rank based on CP decomposition [4,44], the Tucker rank based on Tucker decomposition [8,24,23,46], and the tensor tubal rank based on tensor singular value decomposition (t-SVD) [19,14,43].

The CP rank and the Tucker rank are the two most typical definitions of the tensor rank. The CP rank is defined as the minimum number of rank-one tensors required to express a tensor [20], i.e.,

$$\text{rank}_{\text{cp}}(\mathcal{X}) := \min\left\{ r | \mathcal{X} = \sum_{i=1}^{r} \mathbf{a}_i^1 \circ \mathbf{a}_i^2 \circ \cdots \circ \mathbf{a}_i^N, \mathbf{a}_i^k \in \mathbb{R}^{n_k} \right\}, \tag{2}$$

where $\mathcal{X}$ is an N-way tensor and $\circ$ denotes the vector outer product. Although the measure of the CP rank is consistent with that of the matrix rank, it is difficult to establish a solvable relaxation form. The Tucker rank is defined as a vector, the k-th element of which is the rank of the mode-k unfolding matrix [20], i.e.,

$$\text{rank}_{\text{tc}}(\mathcal{X}) := \left(\text{rank}(X_{(1)}), \text{rank}(X_{(2)}), \cdots, \text{rank}(X_{(N)})\right), \tag{3}$$

where $\mathcal{X}$ is an N-way tensor and $X_{(k)} (k = 1, 2, \cdots, N)$ is the mode-k unfolding of $\mathcal{X}$. To efficiently minimize the Tucker rank, Liu et al. [24] considered its convex relaxation, defined as the sum of the nuclear norm (SNN) of unfolding matrices, i.e.,

$$\|\mathcal{X}\|_{\text{SNN}} := \sum_{k=1}^{N} \alpha_k \|X_{(k)}\|_*, \tag{4}$$

where $\alpha_k \geqslant 0 (k = 1, 2, \cdots, N)$ and $\sum_{k=1}^{N} \alpha_k = 1$. Based on the SNN, Liu et al. [24] established an LRTC model with three solving algorithms (SiLRTC, FaLRTC, and HaLRTC), and Goldfarb and Qin [9] proposed a TRPCA model. Although the SNN can flexibly exploit the correlations along different modes by adjusting the weights α_k [29], as noted in [19,35], when a tensor is unfolded to a matrix along one mode, the structure information along other modes is inevitably destroyed. Thus, the SNN faces difficulty in preserving the intrinsic structure of the tensor. Moreover, Mu et al. [28] showed that the SNN based on standard mode-k unfolding is substantially suboptimal and subsequently offered a generalized tensor unfolding to unfold an N-way tensor to a more balanced (square) matrix, leading to promising results.

Recently, the tensor tubal rank and multi-rank, based on t-SVD, have received considerable attention [19,43,14,16,12,48,25,36,31]. As a generalization of the matrix singular value decomposition (SVD), t-SVD regards a three-way tensor $\mathcal{X}$ as a matrix, each element of which is a tube (mode-3 fiber), and then decomposes $\mathcal{X}$ as

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\text{T}}, \tag{5}$$

where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors, $\mathcal{S}$ is an f-diagonal tensor, $\mathcal{V}^{\text{T}}$ denotes the conjugate transpose of $\mathcal{V}$, and $*$ denotes the t-product (see details in Section 2). Mathematically, this decomposition is equivalent to a series of matrix SVDs in the Fourier domain [43], i.e.,

$$\overline{X}^{(i)} = \overline{U}^{(i)} \overline{S}^{(i)} \left(\overline{V}^{(i)}\right)^{\text{T}}, \quad i = 1, 2, \cdots, n_3, \tag{6}$$

where $\overline{X}^{(i)}$ is the i-th frontal slice of $\overline{\mathcal{X}}$. $\overline{\mathcal{X}}$ is generated by performing the discrete Fourier transformation (DFT) along each tube of $\mathcal{X}$. The multi-rank of $\mathcal{X}$ is defined as a vector whose i-th element is the rank of $\overline{X}^{(i)}$, i.e.,

$$\text{rank}_{\text{m}}(\mathcal{X}) := \left(\text{rank}(\overline{X}^{(1)}), \text{rank}(\overline{X}^{(2)}), \cdots, \text{rank}(\overline{X}^{(n_3)})\right). \tag{7}$$

The tubal rank of $\mathcal{X}$ is defined as the number of non-zero tubes of $\mathcal{S}$, i.e.,

$$\text{rank}_{\text{t}}(\mathcal{X}) := \#\{i : \mathcal{S}(i, i, :) \neq 0\}. \tag{8}$$

Specifically, the tensor tubal rank is equal to the maximum value of the tensor multi-rank. Since directly minimizing the tensor tubal/multi-rank is NP-hard [11], Semerci et al. [31] developed the tensor nuclear norm (TNN) as their convex surrogate, i.e.,

$$\|\mathcal{X}\|_{\text{TNN}} := \sum_{i=1}^{n_3} \|\overline{X}^{(i)}\|_*. \tag{9}$$

Then, Zhang et al. [43] proposed the TNN-based LRTC model, Lu et al. [25] further proved the exactly-recover-property for the TNN-based TRPCA model, and Hu et al. [12] proposed a twist tensor nuclear norm (t-TNN) for video completion.

Although the TNN has shown its effectiveness in preserving the intrinsic structure of a tensor [43,12], it has two obvious shortcomings. One is that it cannot be applied to N-way tensors ($N > 3$). The other is that it lacks the flexibility necessary to address different correlations along different modes, especially the third mode. Specifically, under the framework of t-SVD, for a three-way tensor, the correlations along the first and second modes are characterized by matrix SVD, while that along the third mode is encoded by an embedded circular convolution [43,25]. However, most real-world data always have different correlations along different modes, e.g., the correlation of an HSI along its spectral mode should be much stronger than those along its spatial modes. Thus, treating each mode flexibly similar to the SNN is expected to compensate for this defect.

To apply t-SVD to N-way tensors ($N \geqslant 3$), in this paper, we define a three-way extension of the tensor matricization operator, named mode-k_1k_2 tensor unfolding ($k_1 < k_2$), as the process of lexicographically stacking the mode-k_1k_2 slices of an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ into the frontal slices of a three-way tensor $\mathcal{X}_{(k_1k_2)} \in \mathbb{R}^{n_{k_1} \times n_{k_2} \times \prod_{s \neq k_1,k_2} n_s}$ (see details in Section 3).

To characterize the correlations along different modes in a more flexible manner, we propose a new tensor rank, named the tensor N-tubal rank, which is a vector consisting of the tubal ranks of all mode-k_1k_2 unfolding tensors, i.e.,

$$\begin{aligned} N-\text{rank}_{\text{t}}(\mathcal{X}) := \big(&\text{rank}_{\text{t}}(\mathcal{X}_{(12)}), \text{rank}_{\text{t}}(\mathcal{X}_{(13)}), \cdots, \text{rank}_{\text{t}}(\mathcal{X}_{(1N)}), \\ &\text{rank}_{\text{t}}(\mathcal{X}_{(23)}), \cdots, \text{rank}_{\text{t}}(\mathcal{X}_{(2N)}), \cdots, \text{rank}_{\text{t}}(\mathcal{X}_{(N-1N)})\big) \in \mathbb{R}^{N(N-1)/2}. \end{aligned} \tag{10}$$

Table 1 compares the Tucker rank and the N-tubal rank of two HSIs.[1] As observed, the Tucker rank suggests a strong correlation along the third mode. According to the tensor N-tubal rank, this strong correlation is inadequately depicted by the first element (the tubal rank), while it can be exactly depicted by the other two elements. This observation demonstrates that compared with the tensor tubal rank, the proposed tensor N-tubal rank achieves a more flexible depiction for the correlations along different modes.

To efficiently minimize the proposed tensor N-tubal rank, we establish its convex relaxation: the weighted sum of the tensor nuclear norm (WSTNN), which can be expressed as the weighted sum of the TNN of each mode-k_1k_2 unfolding tensor, i.e.,

$$\|\mathcal{X}\|_{\text{WSTNN}} := \sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1k_2} \|\mathcal{X}_{(k_1k_2)}\|_{\text{TNN}}, \tag{11}$$

where $\alpha_{k_1k_2} \geqslant 0 (1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ and $\sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1k_2} = 1$. Then, we apply the WSTNN to two typical LRTR problems, i.e., LRTC and TRPCA, and propose the corresponding WSTNN-based models. Meanwhile, two efficient alternating direction method of multipliers (ADMM)-based algorithms are developed to solve the proposed models. Numerous numerical experiments on synthetic and real-world data are conducted to illustrate the effectiveness and efficiency of the proposed methods.

The rest of this paper is organized as follows. Section 2 presents some preliminary knowledge. Section 3 gives the definitions of the tensor N-tubal rank and its convex surrogate WSTNN. Section 4 proposes the WSTNN-based LRTC and TRPCA models and develops two efficient ADMM-based solvers. Section 5 evaluates the performance of the proposed models and compares the results with those of state-of-the-art competing methods. Section 6 concludes this paper.

2. Notations and preliminaries

In this section, we give some basic notations and briefly introduce some definitions used throughout the paper [20,43].

We denote vectors as bold lowercase letters (e.g., $\mathbf{x}$), matrices as uppercase letters (e.g., X), and tensors as calligraphic letters (e.g., $\mathcal{X}$). Taking a three-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as an example, we denote its (i,j,s)-th element as $\mathcal{X}(i,j,s)$ or $\mathcal{X}_{i,j,s}$ and its (i,j)-th mode-1, mode-2, and mode-3 fibers as $\mathcal{X}(:,i,j), \mathcal{X}(i,:,j)$, and $\mathcal{X}(i,j,:)$, respectively. We use $\mathcal{X}(i,:,:), \mathcal{X}(:,i,:)$, and $\mathcal{X}(:,:,i)$ to denote the i-th horizontal, lateral, and frontal slices of $\mathcal{X}$, respectively. More compactly, $X^{(i)}$ is short for $\mathcal{X}(:,:,i)$. The Frobenius norm of $\mathcal{X}$ is defined as $\|\mathcal{X}\|_F := \left(\sum_{i,j,s} |\mathcal{X}(i,j,s)|^2\right)^{1/2}$. The ℓ_1 norm of $\mathcal{X}$ is defined as $\|\mathcal{X}\|_1 := \sum_{i,j,s} |\mathcal{X}(i,j,s)|$. We use $\bar{\mathcal{X}}$ to denote the tensor generated by performing DFT along each tube of $\mathcal{X}$, i.e., $\overline{\mathcal{X}} = \texttt{fft}(\mathcal{X}, [], 3)$. Naturally, we can compute $\mathcal{X}$ via $\mathcal{X} = \texttt{ifft}(\overline{\mathcal{X}}, [], 3)$.

The vectorization of an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, denoted as $\mathbf{x} = \texttt{vec}(\mathcal{X}) \in \mathbb{R}^{n_1 n_2 \cdots n_N}$, is defined as

$$\mathbf{x}(j) = \mathcal{X}(i_1, i_2, \cdots, i_N) \text{ with } j = i_1 + \sum_{s=2}^{N} \left((i_s - 1) \prod_{m=1}^{s-1} n_m \right).$$

[1] The rank is approximated by the numbers of singular values larger than 1% of the largest ones.

Table 1
The rank estimation of two HSIs.

Data	Size	Tucker rank	N-tubal rank
Washington DC Mall	$256 \times 256 \times 150$	(107,110,6)	(182,8,8)
Pavia University	$256 \times 256 \times 87$	(115,119,7)	(137,8,8)

The mode-k tensor matricization of an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ is denoted as $X_{(k)} \in \mathbb{R}^{n_k \times \prod_{s \neq k} n_s}$, the (i_k, j)-th element of which maps to the $(i_1, i_2, \cdots, i_N)$-th element of $\mathcal{X}$, where

$$j = 1 + \sum_{s=1, s \neq k}^{N} (i_s - 1) J_s \text{with} J_s = \prod_{m=1, m \neq k}^{s-1} n_m.$$

The corresponding operator and inverse operator are denoted as "$\texttt{unfold}$" and "$\texttt{fold}$", respectively, i.e., $X_{(k)} = \texttt{unfold}(\mathcal{X}, k)$ and $\mathcal{X} = \texttt{fold}(X_{(k)}, k)$.

For a three-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the block circulation operation is defined as

$$\texttt{bcirc}(\mathcal{X}) := \begin{pmatrix} X^{(1)} & X^{(n_3)} & \dots & X^{(2)} \\ X^{(2)} & X^{(1)} & \dots & X^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ X^{(n_3)} & X^{(n_3-1)} & \dots & X^{(1)} \end{pmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}.$$

The block diagonalization operation and its inverse operation are defined as

$$\texttt{bdiag}(\mathcal{X}) := \begin{pmatrix} X^{(1)} & & & \\ & X^{(2)} & & \\ & & \ddots & \\ & & & X^{(n_3)} \end{pmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}, \ \texttt{bdfold}(\texttt{bdiag}(\mathcal{X})) := \mathcal{X}.$$

The block vectorization operation and its inverse operation are defined as

$$\texttt{bvec}(\mathcal{X}) := \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(n_3)} \end{pmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2}, \ \texttt{bvfold}(\texttt{bvec}(\mathcal{X})) := \mathcal{X}.$$

Definition 1 (*t-product*). *The t-product between two three-way tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is defined as*

$$\mathcal{X} * \mathcal{Y} := \texttt{bvfold}(\texttt{bcirc}(\mathcal{X})\texttt{bvec}(\mathcal{Y})) \in \mathbb{R}^{n_1 \times n_4 \times n_3}.$$

Indeed, the t-product can be regarded as a matrix–matrix multiplication, except that the multiplication operation between scalars is replaced by circular convolution between the tubes, i.e.,

$$\mathcal{F} = \mathcal{X} * \mathcal{Y} \Longleftrightarrow \mathcal{F}(i, j, :) = \sum_{t=1}^{n_2} \mathcal{X}(i, t, :) \star \mathcal{Y}(t, j, :),$$

where $\star$ denotes the circular convolution between two tubes. Since that circular convolution in the spatial domain is equivalent to multiplication in the Fourier domain, the t-product between two tensors $\mathcal{F} = \mathcal{X} * \mathcal{Y}$ is equivalent to

$$\bar{\mathcal{F}} = \texttt{bdfold}(\texttt{bdiag}(\bar{\mathcal{X}})\texttt{bdiag}(\bar{\mathcal{Y}})).$$

Definition 2 (*special tensors*). *The conjugate transpose of a three-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\mathcal{X}^{\mathrm{T}}$, is the tensor obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n_3. The identity tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a tensor whose first frontal slice is the identity matrix, and other frontal slices are all zeros. A three-way tensor $\mathcal{Q}$ is orthogonal if $\mathcal{Q} * \mathcal{Q}^{\mathrm{T}} = \mathcal{Q}^{\mathrm{T}} * \mathcal{Q} = \mathcal{I}$. A three-way tensor $\mathcal{S}$ is f-diagonal if each of its frontal slices is a diagonal matrix.*

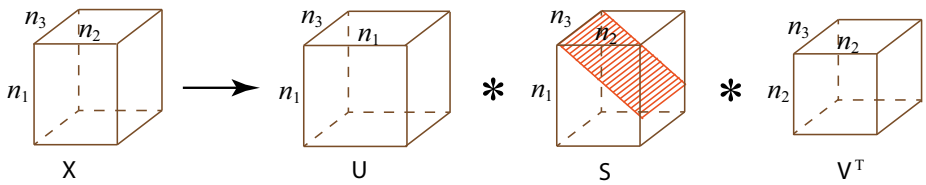

Fig. 1. Illustration of the t-SVD of an $n_1 \times n_2 \times n_3$ tensor.

Algorithm 1 The t-SVD for three-way tensors

Input $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.
1: $\bar{\mathcal{X}} \leftarrow \texttt{fft}(\mathcal{X}, [], 3)$.
2: **for** $i = 1$ to n_3 **do**
3: $[U, S, V] = \texttt{svd}(\overline{X}^{(i)})$.
4: $\overline{U}^{(i)} \leftarrow U; \bar{S}^{(i)} \leftarrow S; \overline{V}^{(i)} \leftarrow V$.
5: **endfor**
6: $\mathcal{U} \leftarrow \texttt{ifft}(\bar{\mathcal{U}}, [], 3)$.
7: $\mathcal{S} \leftarrow \texttt{ifft}(\bar{\mathcal{S}}, [], 3)$.
8: $\mathcal{V} \leftarrow \texttt{ifft}(\bar{\mathcal{V}}, [], 3)$.
Output: $\mathcal{U}$, $\mathcal{S}$, $\mathcal{V}$.

Theorem 1 (*t-SVD*). *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a three-way tensor, then it can be factored as*

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}},$$

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

The t-SVD scheme is illustrated in Fig. 1, and its computation is given in Algorithm 1. Now, we give the definitions of the tensor multi-rank and tubal rank.

Definition 3 (*tensor multi-rank and tubal rank*). *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a three-way tensor. The tensor multi-rank of $\mathcal{X}$ is a vector $\mathrm{rank_m}(\mathcal{X}) \in \mathbb{R}^{n_3}$, the i-th element of which is the rank of the i-th frontal slice of $\bar{\mathcal{X}}$, where $\bar{\mathcal{X}} = \texttt{fft}(\mathcal{X}, [], 3)$. The tubal rank of $\mathcal{X}$, denoted as $\mathrm{rank_t}(\mathcal{X})$, is defined as the number of non-zero tubes of $\mathcal{S}$, where $\mathcal{S}$ comes from the t-SVD of $\mathcal{X}$: $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$. That is, $\mathrm{rank_t}(\mathcal{X}) = \max(\mathrm{rank_m}(\mathcal{X}))$.*

Definition 4 (*tensor nuclear norm (TNN)*). *The tensor nuclear norm of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{X}\|_{\mathrm{TNN}}$, is defined as the sum of the singular values of all the frontal slices of $\bar{\mathcal{X}}$, i.e.,*

$$\|\mathcal{X}\|_{\mathrm{TNN}} := \sum_{i=1}^{n_3} \|\overline{X}^{(i)}\|_*,$$

where $\overline{X}^{(i)}$ is the i-th frontal slice of $\overline{\mathcal{X}}$, with $\bar{\mathcal{X}} = \texttt{fft}(\mathcal{X}, [], 3)$.

3. Tensor *N*-tubal rank and convex relaxation

In this section, we first propose the mode-k_1k_2 tensor unfolding operation and then give the definitions of the tensor *N*-tubal rank and its convex relaxation WSTNN.

As noted in Section 1, the framework of t-SVD and the corresponding tubal rank apply only to three-way tensors and lack the flexibility to handle different correlations along different modes. To address these two issues, we define a novel tensor unfolding operation to transform an *N*-way tensor into a three-way tensor by reordering its slices along any two modes.

Definition 5 (*mode-$\mathtt{k_1k_2}$ slices*). *For an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, its mode-$k_1k_2$ slices ($X^{k_1k_2}, 1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z}$) are two-dimensional sections, defined by fixing all but the mode-k_1 and the mode-k_2 indexes.*

Definition 6 (*mode-$\mathtt{k_1k_2}$ tensor unfolding*). *For an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, its mode-$k_1k_2$ unfolding is a three-way tensor denoted by $\mathcal{X}_{(k_1k_2)} \in \mathbb{R}^{n_{k_1} \times n_{k_2} \times \prod_{s \neq k_1, k_2} n_s}$, the frontal slices of which are the lexicographic orderings of the mode-k_1k_2 slices of $\mathcal{X}$. Mathematically, the $(i_1, i_2, \cdots, i_N)$-th element of $\mathcal{X}$ maps to the (i_{k_1}, i_{k_2}, j)-th element of $\mathcal{X}_{(k_1k_2)}$, where*

$$j = 1 + \sum_{\substack{s=1 \\ s \neq k_1, s \neq k_2}}^{N} (i_s - 1) J_s \text{ with } J_s = \prod_{\substack{m=1 \\ m \neq k_1, m \neq k_2}}^{s-1} n_m.$$

We define the corresponding operation as $\mathcal{X}_{(k_1k_2)} := \mathtt{t-unfold}(\mathcal{X}, k_1, k_2)$ and its inverse operation as $\mathcal{X} := \mathtt{t-fold}(\mathcal{X}_{(k_1k_2)}, k_1, k_2)$. Examples of Definition 5 and Definition 6 can be found in the Appendix. Specifically, for a three-way tensor, the proposed tensor unfolding operation does not involve dimensional reduction but corresponds to a permutation operation, i.e.,

$$\mathcal{X}(i,j,s) = \mathcal{X}_{(12)}(i,j,s) = \mathcal{X}_{(13)}(i,s,j) = \mathcal{X}_{(23)}(j,s,i).$$

Therefore, in this case, we use `permute` and `ipermute` to replace `t-unfold` and `t-fold`, respectively.

By performing t-SVD on each mode-k_1k_2 unfolding tensor, we propose a novel tensor rank, named the tensor N-tubal rank.

Definition 7 (*$\mathtt{N}$-tubal rank*). *NThe-tubal rank of an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ is defined as a vector, the elements of which contain the tubal rank of all mode-k_1k_2 unfolding tensors, i.e.,*

$$N - \mathrm{rank_t}(\mathcal{X}) = \left(\mathrm{rank_t}(\mathcal{X}_{(12)}), \mathrm{rank_t}(\mathcal{X}_{(13)}), \cdots, \mathrm{rank_t}(\mathcal{X}_{(1N)}), \mathrm{rank_t}(\mathcal{X}_{(23)}), \cdots, \mathrm{rank_t}(\mathcal{X}_{(2N)}), \cdots, \mathrm{rank_t}(\mathcal{X}_{(N-1N)})\right) \in \mathbb{R}^{N(N-1)/2}.$$

Clearly, for a three-way tensor, the tensor tubal rank is the first element of the tensor N-tubal rank. By taking the HSI *Washington DC Mall* shown in Fig. 2 as an example, its low N-tubal rank prior can be observed both quantitatively and visually. Specifically, the proposed N-tubal rank combines the advantages of the Tucker rank and tubal rank. On the one hand, compared with the mode-k_1 unfolding matrix, the mode-k_1k_2 unfolding tensor avoids the destruction of the structure information along the k_2-th mode. On the other hand, as shown in Fig. 2, the tubal rank of each mode-k_1k_2 unfolding (permutation) tensor $\mathcal{X}_{(k_1k_2)}$ more directly depicts the correlation of the k_1-th and the k_2-th modes, i.e., it lacks direct characterization of the correlation along other modes. Because all mode-k_1k_2 unfolding tensors are considered simultaneously, the proposed N-tubal rank can effectively exploit the correlations along all modes. The following theorem reveals the relationship between the tensor N-tubal rank and Tucker rank.

Theorem 2 (*$\mathtt{N}$-tubal rank and Tucker rank*). *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ be an N-way tensor with Tucker rank $(r_1, r_2, \cdots, r_N)$ and Tucker decomposition*

$$\mathcal{X} = \mathcal{G} \times_1 A_1 \times_2 A_2 \times_3 \cdots \times_N A_N = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \cdots \sum_{i_N=1}^{r_N} \mathcal{G}(i_1, i_2, \cdots, i_N) \mathbf{a}_{i_1}^1 \circ \mathbf{a}_{i_2}^2 \circ \cdots \circ \mathbf{a}_{i_N}^N,$$

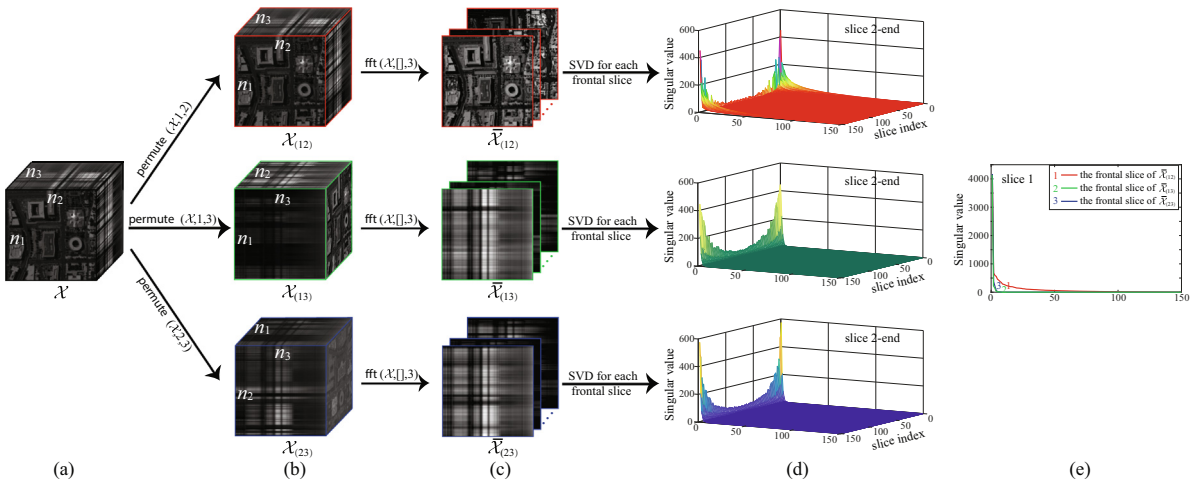

Fig. 2. Illustration of the low N-tubal rank prior of an HSI. (a) The HSI *Washington DC Mall*, which has a size of $150 \times 150 \times 150$. (b) The mode-$k_1k_2$ permutation tensors of $\mathcal{X}$. (c) The tensors $\bar{\mathcal{X}}_{(k_1k_2)}$ generated by performing a DFT along each tube of $\mathcal{X}_{(k_1k_2)}$. (d) Singular value curves from the second to the end frontal slices of $\bar{\mathcal{X}}_{(k_1k_2)}$. (e) Singular value curves of the first frontal slices of $\bar{\mathcal{X}}_{(k_1k_2)}$.

where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_N}, A_k \in \mathbb{R}^{n_k \times r_k} (k = 1, 2, \cdots, N)$, *and* $\mathbf{a}_{i_k}^k$ *is the* i_k*-th column of* A_k. *Then, each element of the N-tubal rank is bounded by the Tucker rank along the corresponding modes, i.e.,*

$$\text{tubal} - \text{Rank}(\mathcal{X}_{(k_1 k_2)}) \leqslant \min\{r_{k_1}, r_{k_2}\}.$$

This theorem demonstrates theoretically that the proposed N-tubal rank learns the global correlations within multi-dimensional data as the Tucker rank does. Furthermore, we reveal the relationship between the tensor N-tubal rank and CP rank in the next theorem.

Theorem 3 (N*-tubal rank and CP rank*). *Assume that the CP rank of an N-way tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ *is r and that its CP decomposition is*

$$\mathcal{X} = \sum_{i=1}^{r} \mathbf{a}_i^1 \circ \mathbf{a}_i^2 \circ \cdots \circ \mathbf{a}_i^N, \mathbf{a}_i^k \in \mathbb{R}^{n_k}, k = 1, 2, \cdots, N.$$

Then, the N-tubal rank of $\mathcal{X}$ *is at most* $r \times ones(N(N-1)/2, 1)$.[2] *Specifically, we define vector sets*

$$\begin{aligned} \mathbb{V}_1 &= \{\mathbf{a}_i^1 | i = 1, 2, \cdots, r\}, \\ \mathbb{V}_2 &= \{\mathbf{a}_i^2 | i = 1, 2, \cdots, r\}, \\ &\vdots \\ \mathbb{V}_N &= \{\mathbf{a}_i^N | i = 1, 2, \cdots, r\}, \end{aligned}$$

and

$$\mathbf{c}_i = \mathrm{vec}(\mathcal{C}_i) \in \mathbb{R}^{\prod_{s \neq k_1, k_2} n_s}, \ i = 1, 2, \cdots, r,$$

where $\mathcal{C}_i = \mathbf{a}_i^1 \circ \mathbf{a}_i^2 \circ \cdots \circ \mathbf{a}_i^{k_1 - 1} \circ \mathbf{a}_i^{k_1 + 1} \circ \cdots \circ \mathbf{a}_i^{k_2 - 1} \circ \mathbf{a}_i^{k_2 + 1} \circ \cdots \circ \mathbf{a}_i^N$. *If each vector set* $\mathbb{V}_i$ *is linearly independent and there is a j such that each j-th element of* $\bar{\mathbf{c}}_i = fft(\mathbf{c}_i)$ *is non-zero, the N-tubal rank is equal to* $r \times ones(N(N-1)/2, 1)$.

Detailed proofs of Theorem 2 and Theorem 3 can be found in the Appendix. To effectively minimize the tensor N-tubal rank, we propose the following WSTNN as its convex relaxation.

Definition 8 (*weighted sum of the tensor nuclear norm*). *The WSTNN of an N-way tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, *denoted as* $\|\mathcal{X}\|_{\text{WSTNN}}$, *is defined as the weighted sum of the TNN of each mode-*$k_1 k_2$ *unfolding tensor, i.e.,*

$$\|\mathcal{X}\|_{\text{WSTNN}} := \sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1 k_2} \|\mathcal{X}_{(k_1 k_2)}\|_{\text{TNN}},$$

where $\alpha_{k_1 k_2} \geqslant 0 (1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ *and* $\sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1 k_2} = 1$.

The weight $\alpha = (\alpha_{11}, \alpha_{12}, \cdots, \alpha_{1N}, \alpha_{23}, \cdots, \alpha_{2N}, \cdots, \alpha_{N-1N})$ is an important parameter for the WSTNN. For the choice of the weight α, we consider the following three cases.

Case 1: The tensor N-tubal rank of the underlying tensor is unknown and cannot be estimated empirically, such as the case of MRI data. Here, the weight α is chosen to be

$$\alpha = \frac{(1, 1, \cdots, 1)}{N(N-1)/2} = \frac{2(1, 1, \cdots, 1)}{N(N-1)}.$$

Case 2: The tensor N-tubal rank of the underlying tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ is known, i.e.,

$$N - \text{rank}_{\text{t}}(\mathcal{X}) = (r_{11}, r_{12}, \cdots, r_{1N}, r_{23}, \cdots, r_{2N}, \cdots, r_{N-1N}).$$

Since $\alpha_{k_1 k_2}$ stands for the contribution of the TNN of the mode-$k_1 k_2$ unfolding tensor $\mathcal{X}_{(k_1 k_2)}$, the value of $\alpha_{k_1 k_2}$ should be dependent on the tubal rank of $\mathcal{X}_{(k_1 k_2)}$ $(r_{k_1 k_2})$ and the size of the first two modes of $\mathcal{X}_{(k_1 k_2)}$ $(n_{k_1} \text{and} n_{k_2})$. Specially, a larger (or smaller) ratio of $r_{k_1 k_2}$ to $\min(n_{k_1}, n_{k_2})$ corresponds to a smaller (or larger) value of $\alpha_{k_1 k_2}$. Therefore, the following strategy is considered to choose the weight α:

$$\alpha_{k_1 k_2} = \frac{e^{\frac{\eta \hat{r}_{k_1 k_2}}{R}}}{\sum_{1 \leqslant k_1 < k_2 \leqslant N} e^{\frac{\eta \hat{r}_{k_1 k_2}}{R}}}, \text{with} R = \sum_{1 \leqslant k_1 < k_2 \leqslant N} \hat{r}_{k_1 k_2}, \ 1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z},$$

where $\hat{r}_{k_1 k_2} = \frac{\min(n_{k_1}, n_{k_2}) - r_{k_1 k_2}}{\min(n_{k_1}, n_{k_2})}$ and η is a balance parameter.

[2] *ones*$(n, 1) \in \mathbb{R}^n$ *is a vector whose elements are all 1.*

Case 3: Particularly for HSIs/MSIs, although their exact N-tubal ranks are unknown, the correlations along their spectral modes should be much stronger than those along their spatial modes. This implies that the value of the first element of the N-tubal rank should be much larger than the values of its second and third elements. Thus, in this case, we empirically choose the weights α as $(\theta, 1, 1)/(2+\theta)$, where θ is a balance parameter.

4. WSTNN-based models and solving algorithms

In this section, we apply the WSTNN to LRTC and TRPCA and propose the WSTNN-based models with ADMM-based solving schemes.

4.1. WSTNN-based LRTC model

Tensor completion aims at estimating the missing elements from an incomplete observation tensor. Considering an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, the proposed WSTNN-based LRTC model is formulated as

$$\begin{aligned} \min_{\mathcal{X}} \quad & \|\mathcal{X}\|_{\text{WSTNN}} \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathcal{X} - \mathcal{F}) = 0, \end{aligned} \tag{12}$$

where $\mathcal{X}$ is the underlying tensor, $\mathcal{F}$ is the observed tensor, Ω is the index set for the known entries, and $\mathcal{P}_\Omega(\mathcal{X})$ is a projection operator that keeps the entries of $\mathcal{X}$ in Ω and sets all others to zero. Let

$$\imath_{\mathbb{S}}(\mathcal{X}) := \begin{cases} 0, & \text{if } \mathcal{X} \in \mathbb{S}, \\ \infty, & \text{otherwise}, \end{cases} \tag{13}$$

where $\mathbb{S} := \{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}, \mathcal{P}_\Omega(\mathcal{X} - \mathcal{F}) = 0\}$. Then (12) can be rewritten as

$$\min_{\mathcal{X}} \sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1 k_2} \|\mathcal{X}_{(k_1 k_2)}\|_{\text{TNN}} + \imath_{\mathbb{S}}(\mathcal{X}), \tag{14}$$

where $\alpha_{k_1 k_2} \geqslant 0 (1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ and $\sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1 k_2} = 1$.

Algorithm 2 ADMM-based optimization algorithm for the proposed WSTNN-based LRTC model (12).

Input: The observed tensor $\mathcal{F}$, index set Ω, weight
$\alpha = (\alpha_{11}, \alpha_{12}, \cdots, \alpha_{1N}, \alpha_{23}, \cdots, \alpha_{2N}, \cdots, \alpha_{N-1N}), \beta = (\beta_{11}, \beta_{12}, \cdots, \beta_{1N}, \beta_{23}, \cdots, \beta_{2N}, \cdots, \beta_{N-1N})$,
$\beta_{\max} = \left(10^{10}, 10^{10}, \cdots, 10^{10}\right)$, and $\gamma = 1.1$.

Initialization: $\mathcal{X}_\Omega^{(0)} = \mathcal{F}_\Omega, \mathcal{X}_{\Omega^c}^{(0)} = 0, \mathcal{Y}_{k_1 k_2}^{(0)} = 0, \mathcal{M}_{k_1 k_2}^{(0)} = 0, p = 0$, and $p_{\max} = 500$. 1: **while** not converged and $p < p_{\max}$ **do**
2: Update $\mathcal{Y}_{k_1 k_2}^{(p+1)}$ via (19), $1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z}$.
3: Update $\mathcal{X}^{(p+1)}$ via (21).
4: Update $\mathcal{M}_{k_1 k_2}^{(p+1)}$ via (17), $1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z}$.
5: $\beta = \min(\gamma\beta, \beta_{\max})$ and $p = p + 1$.
6: **endwhile Output**: The completed tensor $\mathcal{X}$.

Next, we use the ADMM to solve (14). We rewrite (14) as the following equivalent constrained problem

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{Y}_{k_1 k_2}} \quad & \sum_{1 \leqslant k_1 < k_2 \leqslant N} \alpha_{k_1 k_2} \|(\mathcal{Y}_{k_1 k_2})_{(k_1 k_2)}\|_{\text{TNN}} + \imath_{\mathbb{S}}(\mathcal{X}) \\ \text{s.t.} \quad & \mathcal{X} - \mathcal{Y}_{k_1 k_2} = 0, \quad 1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z}. \end{aligned} \tag{15}$$

The augmented Lagrangian function of (15) can be expressed in the following concise form

$$L_{\beta_{k_1 k_2}}(\mathcal{Y}_{k_1 k_2}, \mathcal{X}, \mathcal{M}_{k_1 k_2}) = \sum_{1 \leqslant k_1 < k_2 \leqslant N} \left\{ \alpha_{k_1 k_2} \|(\mathcal{Y}_{k_1 k_2})_{(k_1 k_2)}\|_{\text{TNN}} + \frac{\beta_{k_1 k_2}}{2} g \|\mathcal{X} - \mathcal{Y}_{k_1 k_2} + \frac{\mathcal{M}_{k_1 k_2}}{\beta_{k_1 k_2}} g\|_F^2 \right\} + \imath_{\mathbb{S}}(\mathcal{X}) + \mathcal{C}, \tag{16}$$

where $\mathcal{M}_{k_1 k_2} (1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ are Lagrange multipliers, $\beta_{k_1 k_2}$ $(1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ are penalty parameters, and $\mathcal{C}$ is a variable independent of $\mathcal{X}$ and $\mathcal{Y}_{k_1 k_2}$. Within the framework of the ADMM, $\mathcal{Y}_{k_1 k_2}, \mathcal{X}$, and $\mathcal{M}_{k_1 k_2}$ are alternately updated as

$$\begin{cases} \text{Step1}: \mathcal{Y}_{k_1k_2}^{(p+1)} = \arg\min_{\mathcal{Y}_{k_1k_2}} L_{\beta_{k_1k_2}}\left(\mathcal{Y}_{k_1k_2}, \mathcal{X}^{(p)}, \mathcal{M}_{k_1k_2}^{(p)}\right), \\ \text{Step2}: \mathcal{X}^{(p+1)} = \arg\min_{\mathcal{X}} L_{\beta_{k_1k_2}}\left(\mathcal{Y}_{k_1k_2}^{(p+1)}, \mathcal{X}, \mathcal{M}_{k_1k_2}^{(p)}\right), \\ \text{Step3}: \mathcal{M}_{k_1k_2}^{(p+1)} = \mathcal{M}_{k_1k_2}^{(p)} + \beta_{k_1k_2}\left(\mathcal{X}^{(p+1)} - \mathcal{Y}_{k_1k_2}^{(p+1)}\right). \end{cases} \tag{17}$$

In Step 1, the $\mathcal{Y}_{k_1k_2}(1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ subproblems are

$$\mathcal{Y}_{k_1k_2}^{(p+1)} = \arg\min_{\mathcal{Y}_{k_1k_2}} \alpha_{k_1k_2} \|(\mathcal{Y}_{k_1k_2})_{(k_1k_2)}\|_{\text{TNN}} + \frac{\beta_{k_1k_2}}{2} g \|(\mathcal{X}_{(k_1k_2)})^{(p)} - (\mathcal{Y}_{k_1k_2})_{(k_1k_2)} + \frac{\left((\mathcal{M}_{k_1k_2})_{(k_1k_2)}\right)^{(p)}}{\beta_{k_1k_2}} g\|_F^2. \tag{18}$$

To solve (18), we introduce the following theorem [43].

Theorem 4. *[43] Assuming that $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a three-way tensor, a minimizer to*

$$\min_{\mathcal{Y}} \tau \|\mathcal{Y}\|_{\text{TNN}} + \frac{1}{2}\|\mathcal{Y} - \mathcal{Z}\|_F^2,$$

is given by the tensor singular value thresholding (t-SVT)

$$\mathcal{Y} = \mathcal{D}_\tau(\mathcal{Z}) := \mathcal{U} * \mathcal{S}_\tau * \mathcal{V}^{\text{T}},$$

where $\mathcal{Z} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\text{T}}$ and $\mathcal{S}_\tau$ is an $n_1 \times n_2 \times n_3$ tensor that satisfies

$$\bar{\mathcal{S}}_\tau(i,i,s) = \max(\overline{\mathcal{S}}(i,i,s) - \tau, 0),$$

where $\overline{\mathcal{S}} = fft(\mathcal{S}, [], 3)$ and τ is a threshold.

Via Theorem 4, $\mathcal{Y}_k$ $(1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ can be updated as

$$\mathcal{Y}_{k_1k_2}^{(p+1)} = \texttt{t-fold}\left(\mathcal{D}_{\frac{\tau}{\beta_{k_1k_2}}}\left((\mathcal{X}_{(k_1k_2)})^{(p)} + \frac{\left((\mathcal{M}_{k_1k_2})_{(k_1k_2)}\right)^{(p)}}{\beta_{k_1k_2}}\right), k_1, k_2\right). \tag{19}$$

In Step 2, we solve the following problem

$$\mathcal{X}^{(p+1)} \in \arg\min_{\mathcal{X}} \sum_{1 \leqslant k_1 < k_2 \leqslant N} \frac{\beta_{k_1k_2}}{2} g \|\mathcal{X} - \mathcal{Y}_{k_1k_2}^{(p+1)} + \frac{\mathcal{M}_{k_1k_2}^{(p)}}{\beta_{k_1k_2}} g\|_F^2 + \imath_{\mathbb{S}}(\mathcal{X}), \tag{20}$$

which is differentiable and has a closed-form solution, i.e.,

$$\mathcal{X}^{(p+1)} = \mathcal{P}_{\Omega^c}\left(\frac{\sum_{1 \leqslant k_1 < k_2 \leqslant N} \beta_{k_1k_2}\left(\mathcal{Y}_{k_1k_2}^{(p+1)} - \frac{\mathcal{M}_{k_1k_2}^{(p)}}{\beta_{k_1k_2}}\right)}{\sum_{1 \leqslant k_1 < k_2 \leqslant N} \beta_{k_1k_2}}\right) + \mathcal{P}_\Omega(\mathcal{F}). \tag{21}$$

The pseudocode of the developed algorithm is described in Algorithm 2.

We analyse the computational complexity of the developed algorithm, which involves three subproblems, i.e., the $\mathcal{Y}_{k_1k_2}$ subproblems, the $\mathcal{X}$ subproblem, and the $\mathcal{M}_{k_1k_2}$ subproblems. Updating $\mathcal{Y}_{k_1k_2}$ requires performing SVD on $d_{k_1k_2}$ matrices with a size of (n_{k_1}, n_{k_2}) and fast Fourier transformations (FFT) on $n_{k_1}n_{k_2}$ vectors with a size of $d_{k_1k_2}$, which cost $\mathcal{O}(D[\log(d_{k_1k_2}) + \min(n_{k_1}, n_{k_2})])$, where $D = \prod_{k=1}^N n_k$ and $d_{k_1k_2} = D/(n_{k_1}n_{k_2})$. Updating $\mathcal{X}$ and $\mathcal{M}_{k_1k_2}$ involves only scalar multiplication costing $\mathcal{O}\left(D\sum_{1 \leqslant k_1 < k_2 \leqslant N} 1\right)$. In summary, the computational cost at each iteration is $\mathcal{O}\left(D\sum_{1 \leqslant k_1 < k_2 \leqslant N}[\log(d_{k_1k_2}) + \min(n_{k_1}, n_{k_2})]\right)$.

4.2. WSTNN-based TRPCA model

The TRPCA aims to exactly recover a low-rank tensor corrupted by sparse noise. Considering an N-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, the proposed WSTNN-based TRPCA model can be formulated as

$$\begin{aligned} \min_{\mathcal{L},\mathcal{E}} \quad & \|\mathcal{L}\|_{\text{WSTNN}} + \lambda\|\mathcal{E}\|_1 \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{L} + \mathcal{E}, \end{aligned} \tag{22}$$

where $\mathcal{X}$ is the corrupted observation tensor, $\mathcal{L}$ is the low-rank component, $\mathcal{E}$ is the sparse component, and λ is a tuning parameter compromising $\mathcal{L}$ and $\mathcal{E}$. And (22) can be rewritten as

$$\begin{aligned} \min_{\mathcal{L},\mathcal{E}} \quad & \sum_{1\leqslant k_1<k_2\leqslant N} \alpha_{k_1k_2}\|\mathcal{L}_{(k_1k_2)}\|_{\mathrm{TNN}} + \lambda\|\mathcal{E}\|_1 \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{L}+\mathcal{E}, \end{aligned} \tag{23}$$

where $\alpha_{k_1k_2} \geqslant 0 (1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ and $\sum_{1\leqslant k_1<k_2\leqslant N}\alpha_{k_1k_2} = 1$.

Next, we use the ADMM to solve (23). We rewrite (23) as

$$\begin{aligned} \min_{\mathcal{L},\mathcal{E},\mathcal{Z}_{k_1k_2}} \quad & \sum_{1\leqslant k_1<k_2\leqslant N} \alpha_{k_1k_2}\|(\mathcal{Z}_{k_1k_2})_{(k_1k_2)}\|_{\mathrm{TNN}} + \lambda\|\mathcal{E}\|_1 \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{L}+\mathcal{E}, \\ & \mathcal{L}-\mathcal{Z}_{k_1k_2} = 0, 1\leqslant k_1<k_2\leqslant N, k_1,k_2\in\mathbb{Z}. \end{aligned} \tag{24}$$

The augmented Lagrangian function of (24) can be expressed in the following concise form

$$\begin{aligned} L_{\beta_{k_1k_2},\rho}(\mathcal{L},\mathcal{Z}_{k_1k_2},\mathcal{P}_{k_1k_2},\mathcal{E},\mathcal{M}) = & \sum_{1\leqslant k_1<k_2\leqslant N}\Big\{\alpha_{k_1k_2}\|(\mathcal{Z}_{k_1k_2})_{(k_1k_2)}\|_{\mathrm{TNN}} \\ & + \tfrac{\beta_{k_1k_2}}{2}g\|\mathcal{L}-\mathcal{Z}_{k_1k_2}+\tfrac{\mathcal{P}_{k_1k_2}}{\beta_{k_1k_2}}g\|_F^2\Big\} + \lambda\|\mathcal{E}\|_1 + \tfrac{\rho}{2}g\|\mathcal{X}-\mathcal{L}-\mathcal{E}+\tfrac{\mathcal{M}}{\rho}g\|_F^2 + \mathcal{C}, \end{aligned} \tag{25}$$

where $\mathcal{P}_{k_1k_2}$ and $\mathcal{M}$ are Lagrange multipliers, $\beta_{k_1k_2}$ and ρ are penalty parameters, and $\mathcal{C}$ is a variable independent of $\mathcal{L}, \mathcal{E}$, and $\mathcal{Z}_{k_1k_2}$. To minimize (25), we can update $\mathcal{L}, \mathcal{Z}_{k_1k_2}, \mathcal{P}_{k_1k_2}, \mathcal{E}, \mathcal{M}$ $(1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ as

$$\begin{cases} \text{Step 1}: \mathcal{Z}_{k_1k_2}^{(p+1)} = \arg\min_{\mathcal{Z}_{k_1k_2}} L_{\beta_{k_1k_2},\rho}\Big(\mathcal{L}^{(p)},\mathcal{Z}_{k_1k_2},\mathcal{P}_{k_1k_2}^{(p)},\mathcal{E}^{(p)},\mathcal{M}^{(p)}\Big), \\ \text{Step 2}: \mathcal{L}^{(p+1)} = \arg\min_{\mathcal{L}} L_{\beta_{k_1k_2},\rho}\Big(\mathcal{L},\mathcal{Z}_{k_1k_2}^{(p+1)},\mathcal{P}_{k_1k_2}^{(p)},\mathcal{E}^{(p)},\mathcal{M}^{(p)}\Big), \\ \text{Step 3}: \mathcal{E}^{(p+1)} = \arg\min_{\mathcal{E}} L_{\beta_{k_1k_2},\rho}\Big(\mathcal{L}^{(p+1)},\mathcal{Z}_{k_1k_2}^{(p+1)},\mathcal{P}_{k_1k_2}^{(p)},\mathcal{E},\mathcal{M}^{(p)}\Big), \\ \text{Step 4}: \mathcal{P}_{k_1k_2}^{(p+1)} = \mathcal{P}_{k_1k_2}^{(p)} + \beta_{k_1k_2}\Big(\mathcal{L}^{(p+1)} - \mathcal{Z}_{k_1k_2}^{(p+1)}\Big), \\ \text{Step 5}: \mathcal{M}^{(p+1)} = \mathcal{M}^{(p)} + \rho\big(\mathcal{X}-\mathcal{L}^{(p+1)}-\mathcal{E}^{(p+1)}\big). \end{cases} \tag{26}$$

In Step 1, the $\mathcal{Z}_{k_1k_2}$ $(1 \leqslant k_1 < k_2 \leqslant N, k_1, k_2 \in \mathbb{Z})$ subproblem can be solved as

$$\mathcal{Z}_{k_1k_2}^{(p+1)} = \mathtt{t-fold}\left(\mathcal{D}_{\frac{\alpha_{k_1k_2}}{\beta_{k_1k_2}}}\left(\big(\mathcal{L}_{(k_1k_2)}\big)^{(p)} + \frac{\Big((\mathcal{P}_{k_1k_2})_{(k_1k_2)}\Big)^{(p)}}{\beta_{k_1k_2}}\right), k_1, k_2\right). \tag{27}$$

In Step 2, the $\mathcal{L}$ subproblem has the following closed-form solution

Algorithm 3 ADMM-based optimization algorithm for the proposed WSTNN-based TRPCA model (22).

Input: The corrupted observation tensor $\mathcal{X}$, weight $\alpha = (\alpha_{11}, \alpha_{12}, \cdots, \alpha_{1N}, \alpha_{23}, \cdots, \alpha_{2N}, \cdots, \alpha_{N-1N})$, $\beta = (\beta_{11}, \beta_{12}, \cdots, \beta_{1N}, \beta_{23}, \cdots, \beta_{2N}, \cdots, \beta_{N-1N})$, $\beta_{\max} = \Big(10^{10}, 10^{10}, \cdots, 10^{10}\Big)$, $\lambda, \rho, \rho_{\max} = 10^{10}$, and $\gamma = 1.2$.

1: Initialization: $\mathcal{L}^{(0)} = 0, \mathcal{E}^{(0)} = 0, \mathcal{M}^{(0)} = 0, \mathcal{Z}_{k_1k_2}^{(0)} = 0, \mathcal{P}_{k_1k_2}^{(0)} = 0$, and $p_{\max} = 500$. 2: **while** not converged and $p < p_{\max}$ **do**

3: Update $\mathcal{Z}_{k_1k_2}^{(p+1)}$ via (27), $1 \leqslant k_1 < k_2 \leqslant N$.

4: Update $\mathcal{L}^{(p+1)}$ via (28).

5: Update $\mathcal{E}^{(p+1)}$ via (30).

6: Update $\mathcal{P}_{k_1k_2}^{(p+1)}$ via (26), $1 \leqslant k_1 < k_2 \leqslant N$.

7: Update $\mathcal{M}^{(p+1)}$ via (26).

8: $\beta = \min(\gamma\beta, \beta_{\max}), \rho = \min(\gamma\rho, \rho_{\max})$, and $p = p+1$.

9: **endwhile**

Output: The low-rank component $\mathcal{L}$ and the sparse component $\mathcal{E}$.

$$\mathcal{L}^{(p+1)} = \frac{\rho\left(\mathcal{X} - \mathcal{E}^{(p)} + \frac{\mathcal{M}^{(p)}}{\rho}\right) + \sum\limits_{1\leqslant k_1<k_2\leqslant N} \beta_{k_1k_2}\left(\mathcal{Z}_{k_1k_2}^{(p+1)} - \frac{\mathcal{P}_{k_1k_2}^{(p)}}{\beta_{k_1k_2}}\right)}{\rho + \sum\limits_{1\leqslant k_1<k_2\leqslant N} \beta_{k_1k_2}}. \tag{28}$$

In Step 3, we solve the following problem

$$\mathcal{E}^{(p+1)} \in \arg\min_{\mathcal{E}} \lambda\|\mathcal{E}\|_1 + \frac{\rho}{2} g\|\mathcal{X} - \mathcal{L}^{(p+1)} - \mathcal{E} + \frac{\mathcal{M}^{(p)}}{\rho} g\|_F^2, \tag{29}$$

which has the following closed-form solution

$$\mathcal{E}^{(p+1)} = \mathcal{S}_{\frac{\lambda}{\rho}}\left(\mathcal{X} - \mathcal{L}^{(p+1)} + \frac{\mathcal{M}^{(p)}}{\rho}\right), \tag{30}$$

where $\mathcal{S}_\xi(\cdot)$ is the tensor soft thresholding operator with threshold ξ, i.e.,

$$[\mathcal{S}_\xi(\mathcal{X})]_{i_1i_2\cdots i_N} = \mathrm{sgn}(x_{i_1i_2\cdots i_N}) \max(|x_{i_1i_2\cdots i_N}| - \xi, 0). \tag{31}$$

The pseudocode of the proposed algorithm for solving the proposed WSTNN-based TRPCA model (22) is described in Algorithm 3.

We analyse the detailed computational complexity of the developed algorithm, which involves five subproblems, i.e., the $\mathcal{Z}_{k_1k_2}$ subproblems, the $\mathcal{L}$ subproblem, the $\mathcal{E}$ subproblem, the $\mathcal{P}_{k_1k_2}$ subproblem, and the $\mathcal{M}$ subproblems. Updating $\mathcal{Z}_{k_1k_2}$ requires performing SVD on $d_{k_1k_2}$ matrices with a size of (n_{k_1}, n_{k_2}) and FFT on $n_{k_1}n_{k_2}$ vectors with a size of $d_{k_1k_2}$, which cost $\mathcal{O}(D[\log(d_{k_1k_2}) + \min(n_{k_1}, n_{k_2})])$, where $D = \prod_{k=1}^N n_k$ and $d_{k_1k_2} = D/(n_{k_1}n_{k_2})$. Updating $\mathcal{L}, \mathcal{E}, \mathcal{P}_{k_1k_2}$, and $\mathcal{M}$ involves only scalar multiplication costing $\mathcal{O}\left(D\sum_{1\leqslant k_1<k_2\leqslant N}1\right)$. In summary, the computational cost at each iteration is $\mathcal{O}\left(D\sum_{1\leqslant k_1<k_2\leqslant N}[\log(d_{k_1k_2}) + \min(n_{k_1}, n_{k_2})]\right)$.

5. Numerical experiments

We evaluate the performance of the proposed WSTNN-based LRTC and TRPCA methods.[3] Both synthetic and real-world data are tested. We employ the peak signal-to-noise rate (PSNR), the structural similarity (SSIM) [33], and the feature similarity (FSIM) [41] to measure the quality of the recovered results. All tests are implemented on the Windows 7 platform and MATLAB (R2017b) with an Intel Core i5-4590 3.30 GHz and 16 GB of RAM.

5.1. Low-rank tensor completion

In this section, we test synthetic data and five kinds of real-world data: MSI, HSI, MRI, color video (CV), and hyperspectral video (HSV). If not specified, the methodology for sampling the data is purely random sampling. The compared LRTC methods are as follows: HaLRTC [24] and LRTC-TVI [23], representing the state of the art for the Tucker-decomposition-based method; BCPF [44], representing the state of the art for the CP-decomposition-based method; and logDet [14], TNN [43], PSTNN [16], and t-TNN [12], representing the state of the art for the t-SVD-based method. Because logDet, the TNN, the PSTNN, and the t-TNN apply only to three-way tensors, in all four-way tensor tests, we first reshape the four-way tensors into three-way tensors and then test the performances of these methods.

Parameter selection. In all tests, the stopping criterion depends on the relative change (RelCha) in two successive recovered tensors, i.e., RelCha $= \frac{\|\mathcal{X}^{(p+1)} - \mathcal{X}^{(p)}\|_F}{\|\mathcal{X}^{(p)}\|_F} < 10^{-4}$. Letting the threshold parameter $\tau = \alpha./\beta, \alpha$ is chosen by the weight selection strategy presented in Section 3, τ is set to $\omega \times \texttt{ones}(N(N-1)/2, 1)2$, and ω is empirically selected from a candidate set: $\{1, 10, 50, 100, 500, 1000, 10000\}$. Table 2 shows the parameter settings for the proposed WSTNN-based LRTC method on different data.

Synthetic data completion. We test both synthetic three-way tensors of size $30 \times 30 \times 30$ and four-way tensors of size $30 \times 30 \times 30 \times 30$. The tested synthetic tensors consist of the sum of r rank-one tensors, which are generated by finding the vector outer product on N ($N = 3$ or 4) random vectors. In practice, the data in each test are regenerated and confirmed to meet the conditions of Theorem 3, i.e., the N-tubal rank is $r \times \texttt{ones}(N(N-1)/2, 1)$. We define the success rate as the ratio of successful times to the total number of times, where one test is successful if the relative square error of the recovered tensor $\hat{\mathcal{X}}$ and the ground-truth tensor $\mathcal{X}$, i.e., $\|\hat{\mathcal{X}} - \mathcal{X}\|_F^2/\|\mathcal{X}\|_F^2$, is less than 10^{-3}.

We test data with different N-tubal ranks and sampling rates (SRs), which is defined as the proportion of the known elements. The N-tubal ranks are set to $r \times \texttt{ones}(N(N-1)/2, 1)(r = 1, 2, \ldots, 20)$, and the SRs are set to $0.05 \times s(s = 1, 2, \cdots, 19)$. For each N-tubal rank and SR pair, we conduct 50 independent tests and calculate the success rate. Fig. 3 shows the success

[3] The codes of the WSTNN-based LRTC and TRPCA methods are available at https://yubangzheng.github.io/.

Table 2
Parameter settings of the proposed WSTNN-based LRTC method on different data.

Test	Data		α	τ
synthetic data completion	three-way tensor		(1,1,1)/3	(10,10,10)
	four-way tensor		(1,1,1,1,1,1)/6	(50,50,50,50,50,50)
real-world data completion	three-way tensor	HSI/MSI	(0.001,1,1)/2.001	(100,100,100)
		MRI	(1,1,1)/3	
	four-way tensor	CV	(1,1,1,1,1,1)/6	(500,500,500,500,500,500)
		HSV		

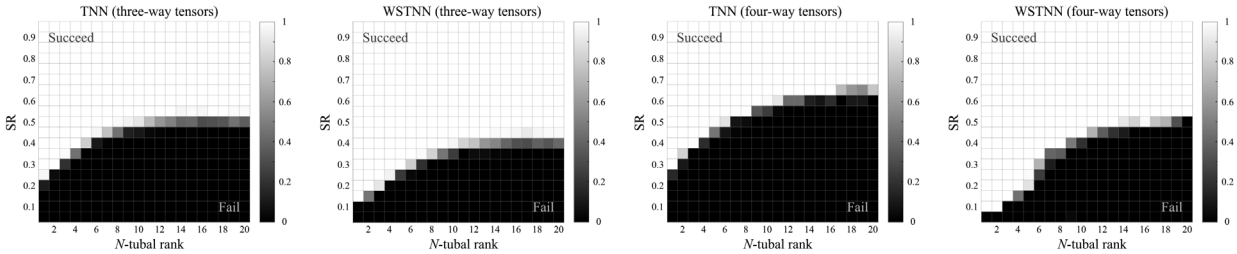

Fig. 3. The success rates for synthetic data with a varying N-tubal rank and varying SR. The left two are the results of the TNN-based LRTC method [43] and the proposed WSTNN-based LRTC method on three-way tensors. The right two are the results of the TNN-based LRTC method [43] and the proposed WSTNN-based LRTC method on four-way tensors. The gray magnitude represents the success rates.

Table 3
The average PSNR, SSIM, and FSIM values for all 32 MSIs tested by the eight utilized LRTC methods.

SR	5%			10%			20%			Time (s)
Method	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
HaLRTC	14.90	0.242	0.648	21.43	0.537	0.773	32.90	0.892	0.933	**13.64**
LRTC-TVI	23.92	0.718	0.812	29.21	0.868	0.895	34.17	0.941	0.953	472.3
BCPF	30.47	0.785	0.884	35.66	0.903	0.936	39.62	0.944	0.962	642.7
logDet	16.99	0.309	0.679	31.27	0.780	0.894	40.81	0.968	0.977	46.31
TNN	17.64	0.332	0.692	30.90	0.780	0.894	39.60	0.962	0.974	46.14
PSTNN	19.56	0.264	0.526	32.95	0.809	0.882	40.77	0.962	0.973	63.48
t-TNN	28.32	0.779	0.874	35.45	0.942	0.954	42.67	0.985	0.987	24.79
WSTNN	**32.03**	**0.881**	**0.930**	**38.74**	**0.977**	**0.979**	**45.70**	**0.994**	**0.994**	75.31

rates for various N-tubal ranks and SRs. It is obvious that under a varying N-tubal rank, the proposed WSTNN-based LRTC method requires less sampling than the TNN-based method [43] to successfully recover the target tensor.

MSI completion. We test 32 MSIs in the dataset CAVE.[4] All testing data are of size $256 \times 256 \times 31$. Table 3 lists the mean values of the PSNR, SSIM, and FSIM for all 32 MSIs recovered by different LRTC methods. As observed, the proposed method can significantly outperform the compared methods in terms of all evaluation indices. To illustrate the visual quality, in Fig. 4, we show one band in three tested data recovered by different methods with SR $= 10\%$. The proposed method is evidently superior to the compared ones in the recovery of both abundant shape structure and texture information. The HSI completion results can be found in the Appendix.

MRI completion. We test an MRI[5] data set of size $181 \times 217 \times 181$. Table 4 lists the values of the PSNR, SSIM, and FSIM of the tested MRI recovered by the different LRTC methods. As observed, the proposed method significantly outperforms the compared methods in terms of all evaluation indices. In Fig. 5, we show three slices obtained in different directions. It can be observed that no matter which direction they are in, the proposed method is evidently superior to the compared ones in the recovery of both abundant shape structure and texture information.

CV completion. We test the CV *news*[6] of size $144 \times 176 \times 3 \times 50$. For each frame, the missing elements of each channel have the same location. Table 5 lists the values of the PSNR, SSIM, and FSIM of the tested CV recovered by different LRTC methods. As

[4] http://www.cs.columbia.edu/CAVE/databases/multispectral.
[5] http://brainweb.bic.mni.mcgill.ca/brainweb/selection_normal.html.
[6] http://trace.eas.asu.edu/yuv/.

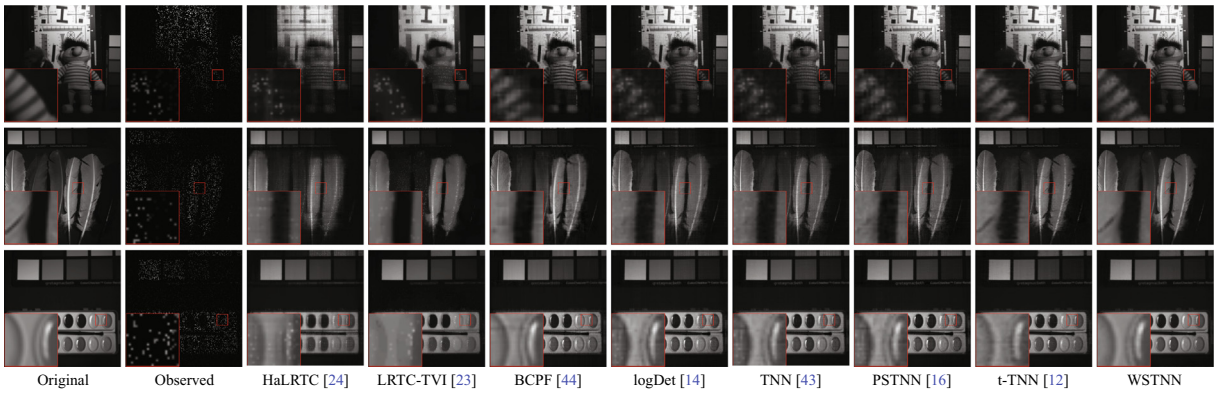

Fig. 4. The completion results of three selected MSIs with SR = 10%. From top to bottom: the images located at the 31-st band in *chart and stuffed toy*, *feathers*, and *paints*, respectively.

Table 4
The PSNR, SSIM, and FSIM values output by the eight utilized LRTC methods for MRI.

SR	5%			10%			20%			Time (s)
Method	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
HaLRTC	15.40	0.241	0.608	19.03	0.390	0.699	24.30	0.653	0.826	**69.981**
LRTC-TVI	19.36	0.597	0.702	22.84	0.748	0.805	28.19	0.891	0.908	1473.8
BCPF	22.37	0.426	0.734	23.81	0.495	0.758	24.96	0.552	0.779	1525.6
logDet	18.32	0.283	0.654	25.36	0.596	0.791	31.22	0.823	0.892	165.90
TNN	22.71	0.472	0.743	26.06	0.642	0.811	29.99	0.799	0.881	165.85
PSTNN	20.39	0.288	0.629	26.45	0.621	0.802	30.71	0.805	0.885	209.19
t-TNN	22.78	0.460	0.736	26.42	0.649	0.816	30.58	0.816	0.890	170.04
WSTNN	**25.60**	**0.714**	**0.827**	**29.02**	**0.835**	**0.887**	**33.46**	**0.931**	**0.941**	405.01

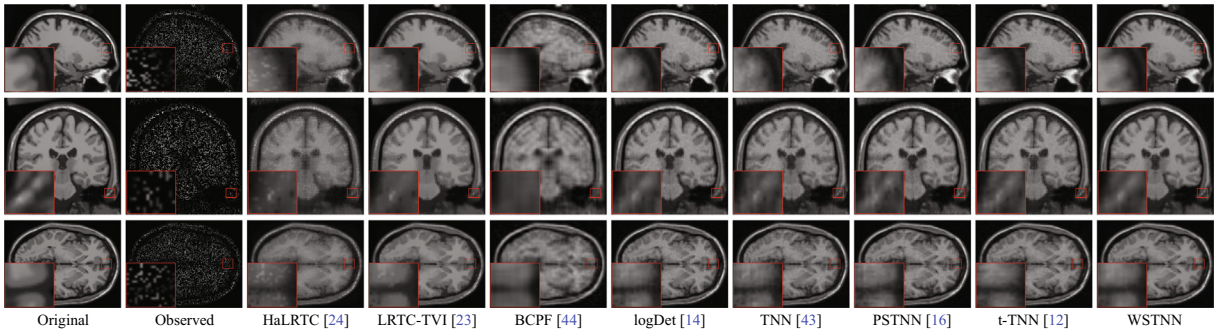

Fig. 5. The completion results of the MRI data with SR = 20%. From top to bottom: the images located at the 70-th horizontal slice, the 100-th lateral slice, and the 70-th frontal slice, respectively.

observed, the proposed method has an overall better performance than that of the compared ones with respect to all evaluation indices. In Fig. 6, we show one frame in the tested CV recovered by the eight compared methods with SR = 10%. We observe that the results obtained by the proposed method are superior to those obtained by the compared ones.

HSV completion. We test an HSV[7] of size $120 \times 120 \times 33 \times 31$. Specifically, this HSV has 31 frames, and each frame has 33 bands of wavelengths of from 400 nm to 720 nm with a 10 nm step [27]. Table 6 lists the values of the PSNR, SSIM, and FSIM of the tested HSV recovered by different LRTC methods. As observed, the proposed method consistently achieves the highest values in terms of all evaluation indexes, e.g., no matter what the SR is set to, the proposed method achieves an approximately 4 dB gain in the PSNR compared with the second-best method. In Fig. 7, we show two images located at different frames and different bands in the HSV recovered by the eight compared methods with SR = 5%. We observe that the proposed method is evidently superior to the compared ones, especially in the recovery of texture information.

[7] http://openremotesensing.net/knowledgebase/hyperspectral-video/.

Table 5
The PSNR, SSIM, and FSIM values output by the eight utilized LRTC methods for CVs.

CV	SR	5%			10%			20%			Time (s)
	Method	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
news											
	HaLRTC	12.59	0.413	0.649	17.67	0.596	0.767	23.92	0.816	0.886	**42.53**
	LRTC-TVI	18.31	0.640	0.731	20.16	0.728	0.802	23.51	0.858	0.901	768.8
	BCPF	25.49	0.779	0.881	28.05	0.857	0.919	29.87	0.897	0.939	961.3
	logDet	13.69	0.288	0.836	18.03	0.534	0.782	33.11	0.944	0.969	92.16
	TNN	21.23	0.659	0.832	29.12	0.893	0.940	32.75	0.943	0.968	97.32
	PSTNN	23.03	0.624	0.884	29.69	0.893	0.942	33.37	0.947	0.970	98.38
	t-TNN	20.65	0.605	0.804	26.92	0.844	0.919	31.91	0.934	0.965	91.36
	WSTNN	**26.92**	**0.892**	**0.929**	**30.67**	**0.947**	**0.964**	**34.61**	**0.976**	**0.983**	324.2

Original Observed HaLRTC [24] LRTC-TVI [23] BCPF [44] logDet [14] TNN [43] PSTNN [16] t-TNN [12] WSTNN

Fig. 6. The completion results at the 49-th frame of the CV *news* with SR = 10%.

Table 6
The PSNR, SSIM, and FSIM values output by the eight utilized LRTC methods for an HSV.

SR	5%			10%			20%			Time (s)
Method	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
HaLRTC	9.008	0.115	0.519	10.46	0.194	0.565	13.41	0.338	0.652	**162.77**
LRTC-TVI	22.09	0.686	0.791	27.08	0.835	0.891	32.19	0.931	0.959	5121.5
BCPF	27.75	0.855	0.907	30.23	0.902	0.934	31.69	0.917	0.945	5840.6
logDet	31.01	0.912	0.948	38.94	0.975	0.984	44.52	0.991	0.995	446.61
TNN	33.68	0.946	0.968	38.02	0.974	0.984	42.94	0.989	0.993	487.95
PSTNN	32.93	0.934	0.960	38.53	0.975	0.985	43.41	0.989	0.994	423.32
t-TNN	29.43	0.894	0.931	34.37	0.957	0.971	40.11	0.986	0.990	391.87
WSTNN	**37.61**	**0.979**	**0.986**	**43.67**	**0.994**	**0.996**	**49.11**	**0.997**	**0.998**	1228.3

5.2. Tensor robust principal component analysis

In this section, we evaluate the performance of the proposed WSTNN-based TRPCA method by synthetic data and HSI denoising. The compared TRPCA methods include the SNN [9] and TNN [25].

Parameter selection. In all tests, the stopping criterion depends on the RelCha in two successive recovered tensors, i.e., RelCha $= \frac{\|\mathcal{L}^{(p+1)}-\mathcal{L}^{(p)}\|_F}{\|\mathcal{L}^{(p)}\|_F} < 10^{-4}$. The tuning parameter λ is set to

$$\lambda = \sum_{1\leqslant k_1<k_2\leqslant N} \frac{\alpha_{k_1k_2}}{\sqrt{\max(n_{k_1},n_{k_2})d_{k_1k_2}}} \text{with} d_{k_1k_2} = \prod_{s\neq k_1,k_2} n_s.$$

Letting the threshold parameter $\tau = \alpha./\beta$, the penalty parameter ρ is set to $\rho = 1/\text{mean}(\tau)$. This means that only the weight α and the threshold τ need to be adjusted. Table 7 shows these two parameter settings for the proposed WSTNN-based TRPCA method on different data, where α is chosen by the weight selection strategy presented in Section 3, τ is set to $\omega \times \texttt{ones}(N(N-1)/2,1)$, and ω is empirically selected from a candidate set: $\{1,10,50,100,500,1000,10000\}$.

Synthetic data denoising. We test three-way tensors of size $30\times30\times30$ and four-way tensors of size $30\times30\times30\times30$ with different N-tubal ranks and random salt-pepper noise levels (NLs). The N-tubal ranks are set to $r\times\texttt{ones}(N(N-1)/2,1)(r=1,2,\ldots,20)$, and the NLs are set to $0.025\times l(l=1,2,\cdots,20)$. For each N-tubal rank and NL pair, we conduct 50 independent tests and calculate the success rate. Fig. 8 shows the success rates for varying N-tubal rank and varying NL. The results illustrate that the proposed WSTNN-based TRPCA method is more robust and preferable than the TNN-based method [25].

HSI denoising. We test the *Washington DC Mall* and *Pavia University* HSI data sets. The random salt-pepper NL is set to 0.2 and 0.4. Table 8 lists the PSNR, SSIM, and FSIM values of the tested HSIs recovered by different methods. From these results, we observe that our method evidently performs better than the other competing ones in terms of all the evaluation mea-

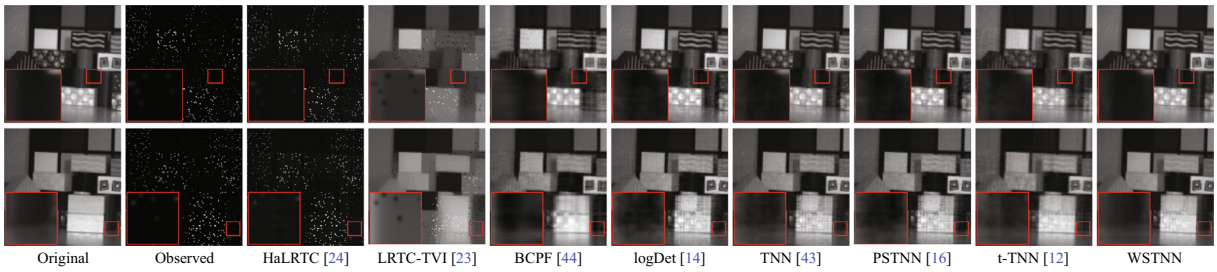

Fig. 7. The completion results of an HSV with SR = 5%. Top row: the image located at the 15-th band and the 7-th frame. Bottom row: the image located at the 25-th band and the 30-th frame.

Table 7
Parameter settings of the proposed WSTNN-based TRPCA method on different data.

Test	Tensor	α	τ
synthetic data denoising	three-way tensor	(1,1,1)/3	(10,10,10)
	four-way tensor	(1,1,1,1,1,1)/6	(50,50,50,50,50,50)
HSI denoising	three-way tensor	(0.001,1,1)/2.001	(100,100,100)

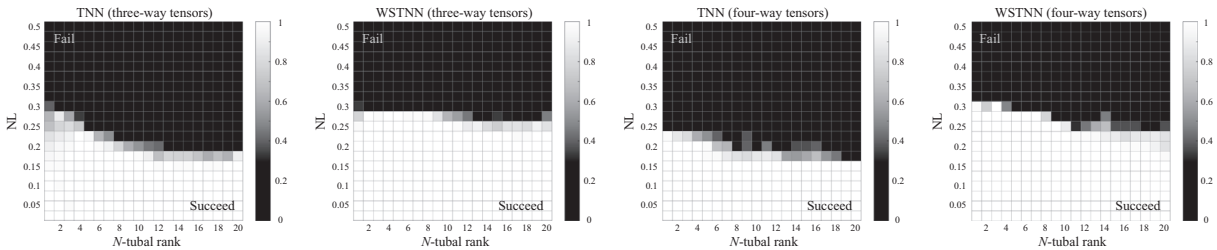

Fig. 8. The success rates for synthetic data with varying *N*-tubal rank and varying NLs. The left two are the results of the TNN-based TRPCA method [25] and the proposed WSTNN-based TRPCA method on three-way tensors. The right two are the results of the TNN-based TRPCA method [25] and the proposed WSTNN-based TRPCA method on four-way tensors. The gray magnitude represents the success rates.

Table 8
The PSNR, SSIM, and FSIM values output by the three utilized TRPCA methods for HSIs.

HSI	NL	0.2			0.4			Time (s)
	Method	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
Washington DC Mall $256 \times 256 \times 150$	SNN	31.48	0.927	0.950	28.19	0.848	0.902	**79.822**
	TNN	43.87	0.992	0.994	35.82	0.953	0.973	172.81
	WSTNN	**50.49**	**0.999**	**0.999**	**42.29**	**0.993**	**0.995**	385.39
Pavia University $256 \times 256 \times 87$	SNN	28.14	0.877	0.899	26.16	0.787	0.834	**56.238**
	TNN	38.97	0.983	0.988	35.42	0.958	0.975	120.28
	WSTNN	**39.21**	**0.995**	**0.997**	**36.48**	**0.988**	**0.993**	243.89

sures. In Fig. 9, we show one band in these two HSIs. As observed, our WSTNN-based TRPCA method achieves the best visual results among those of the three compared methods in terms of both noise removal and detail preservation.

5.3. Parameter study and convergence analysis

In this section, we discuss the effects of the threshold parameter τ and the convergence of the proposed ADMM in the proposed LRTC and TRPCA problems. All tests are based on the HSI *Washington DC Mall*.

Effects of the threshold parameter. We set the SR to 10% in the completion tests and the NL to 0.4 in the denoising tests. In addition, $\tau = (\omega, \omega, \omega)$. The results are presented in Fig. 10(a). As observed, values of τ that are too large or too small result in failure, while moderate values yield the best results. This observation is consistent with the theoretical analysis. That is,

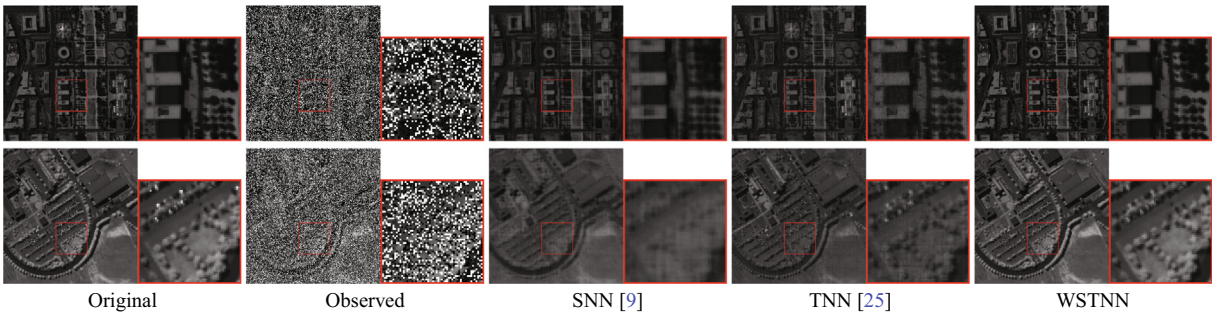

Fig. 9. The denoising results of the HSIs *Washington DC Mall* and *Pavia University* with NL = 0.4. Top row: the image located at the 150-th band in *Washington DC Mall*. Bottom row: the image located at the 87-th band in *Pavia University*.

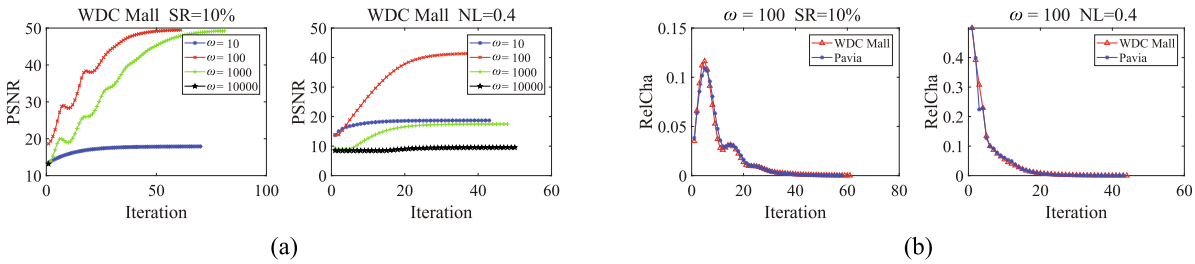

Fig. 10. (a) The PSNR values with respect to the iteration for different values of τ. Left column: completion tests. Right column: denoising tests. (b) The RelCha values with respect to the iteration for $\tau = (100, 100, 100)$. Left column: completion tests. Right column: denoising tests.

for the completion tests, if τ is too large (e.g., $(10000, 10000, 10000)$), all the singular values are replaced with 0, andthe algorithm iterates only one step and outputs the partial observation tensor $\mathcal{F}$. If the parameter τ is too small (e.g., $(10, 10, 10)$), the singular values obtained after performing the t-SVT (in Theorem 4) contain corrupted information, which is not consistent with the low-rank prior of the underlying tensor. Similarly, for the denoising tests, if the parameter τ is too large or too small, the low-rank term becomes out of action. Under the guidance of Fig. 10(a), τ is set to $(100, 100, 100)$ in all experiments conducted on real-world data.

Convergence analysis. Owing to the use of the ADMM framework and the convexity of the objective functions, the convergence of the two developed algorithms is guaranteed theoretically. Empirically, this convergence can be visually observed in Fig. 10(b), where τ is set to $(100, 100, 100)$.

6. Conclusions

In this paper, we defined mode-k_1k_2 tensor unfolding, which is used to reorder the elements of an N-way tensor into a three-way tensor, and then performed t-SVD on each mode-k_1k_2 unfolding tensor to depict the correlations along different modes. On this basis, we proposed the corresponding tensor N-tubal rank and its convex relaxation WSTNN. To illustrate the effectiveness of the proposed N-tubal rank and WSTNN, we applied the WSTNN to two typical LRTR problems, i.e., LRTC and TRPCA problems, and proposed the WSTNN-based LRTC and TRPCA models. Meanwhile, two efficient ADMM-based algorithms were developed to solve the proposed models. The numerical results demonstrated that the proposed method effectively exploits the correlations along all modes while preserving the intrinsic structure of the underlying tensor.

For future work, there are three directions. First, the mechanism of all low-rank models lies in the assumption that the original data has a stronger low-rankness than the observed one. Therefore, the proposed method tends to fail when the observed data have the same, or even stronger, low-N-tubal-rank property compared with the original one. One challenging example is the missing slice problem, which usually results in observed data with a lower N-tubal rank than that of the original data. To solve this issue and further improve the completion performance, we plan to combine the proposed global low-N-tubal-rankness prior to some other priors, such as the piecewise smoothness prior, nonlocal self-similarity prior, and deep prior. Second, we plan to establish some nonconvex relaxations [39,40,42] to further improve the performance of the proposed method. Third, for MSIs/HSIs, we plan to combine the proposed WSTNN with the recent excellent MSI/HSI processing methods, such as FastHyDe [49] and NG-meet [10], to enhance the ability to recover the target HSI.

CRediT authorship contribution statement

Yu-Bang Zheng: Conceptualization, Methodology, Software, Investigation, Data curation, Writing - original draft. **Ting-Zhu Huang:** Conceptualization, Methodology, Project administration, Supervision, Writing - review & editing. **Xi-Le Zhao:** Software, Investigation, Project administration, Supervision, Writing - review & editing. **Tai-Xiang Jiang:** Investigation, Writing - review & editing, Visualization. **Teng-Yu Ji:** Formal analysis, Writing - review & editing, Visualization. **Tian-Hui Ma:** Writing - review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express their sincere thanks to the reviewers for their helpful suggestions regarding how to revise this paper. This work is supported by the National Natural Science Foundation of China (61772003, 61876203, and 11901450), the Fundamental Research Funds for the Central Universities (31020180QD126), the National Postdoctoral Program for Innovative Talents (BX20180252), and the Project funded by China Postdoctoral Science Foundation (2018M643611). The authors would like to express their sincere thanks to the medical staff on the front line of fighting the COVID-19. It is their dedication and sacrifice that provide people a secure and stable research environment in this special time.

Appendix A

1. Examples of Definition 5.

For a four-way tensor $\mathcal{X} \in \mathbb{R}^{2\times3\times3\times2}$, its (i_2, i_4)-th mode-13 slice and (i_1, i_3)-th mode-24 slice are

$$X^{13} = \begin{pmatrix} \mathcal{X}(\mathbf{1}, i_2, \mathbf{1}, i_4) & \mathcal{X}(\mathbf{1}, i_2, \mathbf{2}, i_4) & \mathcal{X}(\mathbf{1}, i_2, \mathbf{3}, i_4) \\ \mathcal{X}(\mathbf{2}, i_2, \mathbf{1}, i_4) & \mathcal{X}(\mathbf{2}, i_2, \mathbf{2}, i_4) & \mathcal{X}(\mathbf{2}, i_2, \mathbf{3}, i_4) \end{pmatrix} \quad \text{and} \quad X^{24} = \begin{pmatrix} \mathcal{X}(i_1, \mathbf{1}, i_3, \mathbf{1}) & \mathcal{X}(i_1, \mathbf{1}, i_3, \mathbf{2}) \\ \mathcal{X}(i_1, \mathbf{2}, i_3, \mathbf{1}) & \mathcal{X}(i_1, \mathbf{2}, i_3, \mathbf{2}) \\ \mathcal{X}(i_1, \mathbf{3}, i_3, \mathbf{1}) & \mathcal{X}(i_1, \mathbf{3}, i_3, \mathbf{2}) \end{pmatrix},$$

respectively.

2. Examples of Definition 6.

For a four-way tensor $\mathcal{X} \in \mathbb{R}^{2\times3\times3\times2}$, its mode-24 unfolding tensor $\mathcal{X}_{(24)} \in \mathbb{R}^{3\times2\times6}$ can be expressed as

$$\mathcal{X}_{(24)}(:,:,1) = \begin{pmatrix} \mathcal{X}(1,\mathbf{1},1,\mathbf{1}) & \mathcal{X}(1,\mathbf{1},1,\mathbf{2}) \\ \mathcal{X}(1,\mathbf{2},1,\mathbf{1}) & \mathcal{X}(1,\mathbf{2},1,\mathbf{2}) \\ \mathcal{X}(1,\mathbf{3},1,\mathbf{1}) & \mathcal{X}(1,\mathbf{3},1,\mathbf{2}) \end{pmatrix}, \quad \mathcal{X}_{(24)}(:,:,4) = \begin{pmatrix} \mathcal{X}(1,\mathbf{1},2,\mathbf{1}) & \mathcal{X}(1,\mathbf{1},2,\mathbf{2}) \\ \mathcal{X}(1,\mathbf{2},2,\mathbf{1}) & \mathcal{X}(1,\mathbf{2},2,\mathbf{2}) \\ \mathcal{X}(1,\mathbf{3},2,\mathbf{1}) & \mathcal{X}(1,\mathbf{3},2,\mathbf{2}) \end{pmatrix},$$

$$\mathcal{X}_{(24)}(:,:,2) = \begin{pmatrix} \mathcal{X}(2,\mathbf{1},1,\mathbf{1}) & \mathcal{X}(2,\mathbf{1},1,\mathbf{2}) \\ \mathcal{X}(2,\mathbf{2},1,\mathbf{1}) & \mathcal{X}(2,\mathbf{2},1,\mathbf{2}) \\ \mathcal{X}(2,\mathbf{3},1,\mathbf{1}) & \mathcal{X}(2,\mathbf{3},1,\mathbf{2}) \end{pmatrix}, \quad \mathcal{X}_{(24)}(:,:,5) = \begin{pmatrix} \mathcal{X}(2,\mathbf{1},2,\mathbf{1}) & \mathcal{X}(2,\mathbf{1},2,\mathbf{2}) \\ \mathcal{X}(2,\mathbf{2},2,\mathbf{1}) & \mathcal{X}(2,\mathbf{2},2,\mathbf{2}) \\ \mathcal{X}(2,\mathbf{3},2,\mathbf{1}) & \mathcal{X}(2,\mathbf{3},2,\mathbf{2}) \end{pmatrix},$$

$$\mathcal{X}_{(24)}(:,:,3) = \begin{pmatrix} \mathcal{X}(3,\mathbf{1},1,\mathbf{1}) & \mathcal{X}(3,\mathbf{1},1,\mathbf{2}) \\ \mathcal{X}(3,\mathbf{2},1,\mathbf{1}) & \mathcal{X}(3,\mathbf{2},1,\mathbf{2}) \\ \mathcal{X}(3,\mathbf{3},1,\mathbf{1}) & \mathcal{X}(3,\mathbf{3},1,\mathbf{2}) \end{pmatrix}, \quad \mathcal{X}_{(24)}(:,:,6) = \begin{pmatrix} \mathcal{X}(3,\mathbf{1},2,\mathbf{1}) & \mathcal{X}(3,\mathbf{1},2,\mathbf{2}) \\ \mathcal{X}(3,\mathbf{2},2,\mathbf{1}) & \mathcal{X}(3,\mathbf{2},2,\mathbf{2}) \\ \mathcal{X}(3,\mathbf{3},2,\mathbf{1}) & \mathcal{X}(3,\mathbf{3},2,\mathbf{2}) \end{pmatrix}.$$

3. Proof of Theorem 2. (N-tubal rank and Tucker rank)

Proof. Apparently, the mode-k_1k_2 unfolding tensor of $\mathcal{X}$ can be expressed as

$$\mathcal{X}_{(k_1k_2)} = \sum_{i_{k_1}=1}^{r_{k_1}} \sum_{i_{k_2}=1}^{r_{k_2}} \mathbf{a}_{i_{k_1}}^{k_1} \circ \mathbf{a}_{i_{k_2}}^{k_2} \circ \mathbf{c}_{i_{k_1} i_{k_2}},$$

where $\mathbf{c}_{i_{k_1} i_{k_2}} = \mathrm{vec}\left(\mathcal{C}_{i_{k_1} i_{k_2}}\right)$ with

Table 9
The PSNR, SSIM, and FSIM values output by the eight utilized LRTC methods for HSIs.

HSI	SR	5%			10%			20%			Time (s)
	Method	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
Washington DC Mall $256 \times 256 \times 150$	HaLRTC	20.72	0.452	0.665	24.74	0.656	0.798	29.38	0.848	0.909	**76.487**
	LRTC-TVI	21.93	0.437	0.605	25.89	0.638	0.759	29.11	0.824	0.893	2348.2
	BCPF	29.07	0.820	0.895	31.89	0.895	0.934	32.77	0.911	0.943	2955.9
	logDet	25.22	0.685	0.848	32.50	0.911	0.947	37.99	0.969	0.981	237.18
	TNN	28.87	0.831	0.907	32.41	0.913	0.949	36.85	0.963	0.977	294.46
	PSTNN	28.15	0.793	0.886	32.63	0.911	0.946	37.39	0.965	0.978	306.16
	t-TNN	33.23	0.932	0.959	43.96	0.994	0.996	56.99	0.997	0.998	184.23
	WSTNN	**40.54**	**0.988**	**0.992**	**50.31**	**0.999**	**0.999**	**58.89**	**0.999**	**0.999**	544.26
Pavia University $256 \times 256 \times 87$	HaLRTC	15.01	0.043	0.517	24.02	0.611	0.736	27.59	0.788	0.861	**49.745**
	LRTC-TVI	23.26	0.554	0.652	25.80	0.713	0.785	29.19	0.866	0.903	1427.3
	BCPF	27.64	0.726	0.835	30.39	0.836	0.898	32.07	0.884	0.928	1603.6
	logDet	26.90	0.684	0.835	32.69	0.876	0.932	39.34	0.959	0.977	140.96
	TNN	28.12	0.750	0.865	32.15	0.874	0.931	37.49	0.950	0.972	168.44
	PSTNN	23.18	0.449	0.737	32.97	0.872	0.932	38.84	0.955	0.974	181.04
	t-TNN	33.38	0.928	0.957	41.15	0.988	0.993	50.83	0.997	0.998	101.49
	WSTNN	**37.26**	**0.976**	**0.983**	**44.48**	**0.995**	**0.997**	**53.92**	**0.999**	**0.999**	258.78

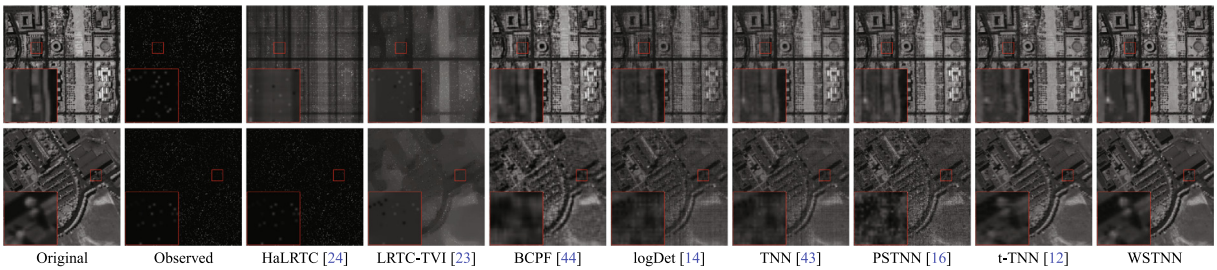

Fig. 11. The completion results of the HSIs *Washington DC Mall* and *Pavia University* with SR = 5%. Top row: the image located at the 70-th band in *Washington DC Mall*. Bottom row: the image located at the 85-th band in *Pavia University*.

$$\mathcal{C}_{i_{k_1} i_{k_2}} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_{k_1-1}=1}^{r_{k_1-1}} \sum_{i_{k_1+1}=1}^{r_{k_1+1}} \cdots \sum_{i_{k_2-1}=1}^{r_{k_2-1}} \sum_{i_{k_2+1}=1}^{r_{k_2+1}} \cdots \sum_{i_N=1}^{r_N} \mathcal{G}(i_1, i_2, \cdots, i_N) \mathbf{a}_{i_1}^1 \circ \cdots \circ \mathbf{a}_{i_{k_1-1}}^{k_1-1} \circ \mathbf{a}_{i_{k_1+1}}^{k_1+1} \circ \cdots \circ \mathbf{a}_{i_{k_2-1}}^{k_2-1} \circ \mathbf{a}_{i_{k_2+1}}^{k_2+1} \circ \cdots \circ \mathbf{a}_{i_N}^N.$$

Letting $\bar{\mathcal{X}}_{(k_1 k_2)} = \texttt{fft}(\mathcal{X}_{(k_1 k_2)}, [], 3)$, then $\bar{\mathcal{X}}_{(k_1 k_2)}$ can be expressed as

$$\bar{\mathcal{X}}_{(k_1 k_2)} = \sum_{i_{k_1}=1}^{r_{k_1}} \sum_{i_{k_2}=1}^{r_{k_2}} \mathbf{a}_{i_{k_1}}^{k_1} \circ \mathbf{a}_{i_{k_2}}^{k_2} \circ \bar{\mathbf{c}}_{i_{k_1} i_{k_2}},$$

where $\bar{\mathbf{c}}_{i_{k_1} i_{k_2}} = \texttt{fft}\left(\mathbf{c}_{i_{k_1} i_{k_2}}\right)$. Letting $\bar{\mathbf{c}}_{i_{k_1} i_{k_2}} = \left(\bar{c}^1_{i_{k_1} i_{k_2}}, \bar{c}^2_{i_{k_1} i_{k_2}} \cdots \bar{c}^d_{i_{k_1} i_{k_2}}\right)^{\mathrm{T}}$ and supposing $r_{k_1} = \min\{r_{k_1}, r_{k_2}\}$, then the j-th $(\forall j = 1, 2, \cdots, d)$ frontal slice of $\bar{\mathcal{X}}_{(k_1 k_2)}$ can be expressed as

$$\overline{X}^{(j)}_{(k_1 k_2)} = \mathbf{a}_1^{k_1} \left(\mathbf{b}_1^j\right)^{\mathrm{T}} + \mathbf{a}_2^{k_1} \left(\mathbf{b}_2^j\right)^{\mathrm{T}} + \cdots + \mathbf{a}_{r_{k_1}}^{k_1} \left(\mathbf{b}_{r_{k_1}}^j\right)^{\mathrm{T}},$$

where $\mathbf{b}_{i_{k_1}}^j = \sum_{i_{k_2}=1}^{r_{k_2}} \bar{c}^j_{i_{k_1} i_{k_2}} \mathbf{a}_{i_{k_2}}^{k_2}$ $(i_{k_1} = 1, 2, \cdots, r_{k_1})$. This implies that the rank of each frontal slice of $\bar{\mathcal{X}}_{(k_1 k_2)}$ is at most r_1. Thus, the theorem holds. □

4. Proof of Theorem 3 (N-tubal rank and CP rank).

Proof. The $\mathcal{X}_{(k_1 k_2)}$ has the following CP decomposition

$$\mathcal{X}_{(k_1 k_2)} = \sum_{i=1}^{r} \mathbf{a}_i^{k_1} \circ \mathbf{a}_i^{k_2} \circ \mathbf{c}_i,$$

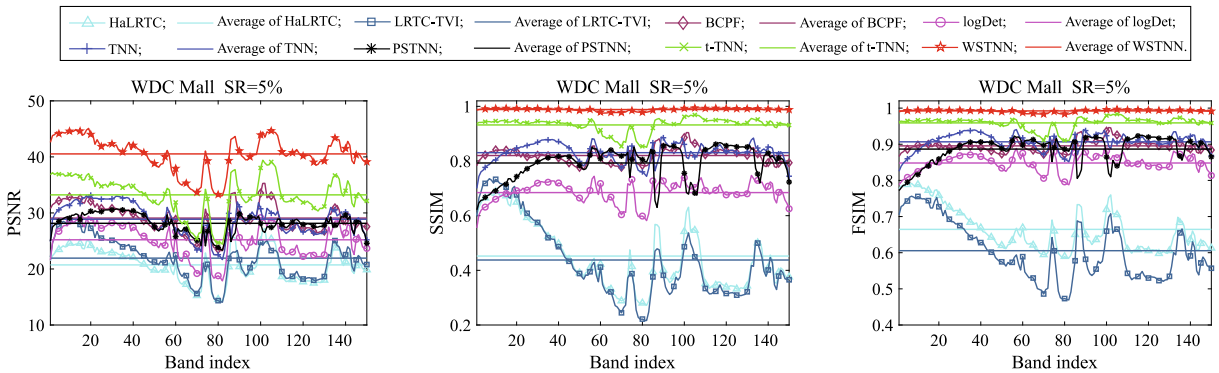

Fig. 12. The PSNR, SSIM, and FSIM values of each band of the recovered HSI *Washington DC Mall* output by the eight LRTC methods with SR = 5%.

Letting $\bar{\mathcal{X}}_{(k_1k_2)} = \texttt{fft}(\mathcal{X}_{(k_1k_2)}, [], 3)$, then $\bar{\mathcal{X}}_{(k_1k_2)}$ has the following CP decomposition

$$\bar{\mathcal{X}}_{(k_1k_2)} = \sum_{i=1}^{r} \mathbf{a}_i^{k_1} \circ \mathbf{a}_i^{k_2} \circ \bar{\mathbf{c}}_i,$$

where $\bar{\mathbf{c}}_i = \texttt{fft}(\mathbf{c}_i)$. Letting $\bar{\mathbf{c}}_i = (\bar{c}_i^1, \bar{c}_i^2 \cdots \bar{c}_i^d)$, then the j-th frontal slice of $\bar{\mathcal{X}}_{(k_1k_2)}$ can be expressed as

$$\bar{X}_{(k_1k_2)}^{(j)} = \bar{c}_1^j \mathbf{a}_1^{k_1} \left(\mathbf{a}_1^{k_2}\right)^{\mathrm{T}} + \bar{c}_2^j \mathbf{a}_2^{k_1} \left(\mathbf{a}_2^{k_2}\right)^{\mathrm{T}} + \cdots + \bar{c}_r^j \mathbf{a}_r^{k_1} \left(\mathbf{a}_r^{k_2}\right)^{\mathrm{T}}.$$

This implies that the rank of each frontal slice of $\bar{\mathcal{X}}_{(k_1k_2)}$ is at most r, and it is equal to r if the vector sets $\mathbb{V}_{k_1}$ or $\mathbb{V}_{k_2}$ is linearly independent and the j-th element of each $\bar{\mathbf{c}}_i$ is non-zero. Thus, the tubal rank of $\mathcal{X}_{(k_1k_2)}$ (the (k_1, k_2)-th element of the N-tubal rank of $\mathcal{X}$) is at most r, and it is equal to r if the aforementioned conditions are satisfied. □

5. HSI completion.

We test HSIs *Washington DC Mall*[8] and *Pavia University*8. Table 9 lists the values of the PSNR, SSIM, and FSIM of these two tested HSIs recovered by different LRTC methods. We observe that compared with other methods, the proposed method consistently achieves the highest values in terms of all evaluation indexes, e.g., when SR is set as 5% or 10%, the proposed method achieves around 7 dB gain in PSNR beyond the second-best method in the test on *Washington DC Mall*. For visual comparison, in Fig. 11, we show one band in these two testing HSIs recovered by the eight utilized LRTC methods with SR = 5%. As observed, the proposed method can produce visually superior results than the compared methods. Fig. 12 shows the PSNR, SSIM and FSIM values of each band of the recovered HSI *Washington DC Mall* obtained by the eight compared LRTC methods with SR = 5%. From this figure, it is easy to observe that the proposed method achieves the best performance in all bands among eight LRTC methods.

References

[1] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (4) (2010) 1956–1982.
[2] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717–772.
[3] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, Z. Xu, Total variation regularized tensor RPCA for background subtraction from compressive measurements, IEEE Trans. Image Process. 25 (9) (2016) 4075–4090.
[4] J.D. Carroll, S. Pruzansky, J.B. Kruskal, Candelinc: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters, Psychometrika 45 (1) (1980) 3–24.
[5] M.A. Davenport, J. Romberg, An overview of low-rank matrix recovery from incomplete observations, IEEE J. Sel. Top. Signal Process. 10 (4) (2016) 608–622.
[6] M. Ding, T.-Z. Huang, T.-Y. Ji, T.-Y. Ji, X.-L. Zhao, J.-H. Yang, Low-rank tensor completion using matrix factorization based on tensor train rank and total variation, J. Sci. Comput. 81 (2) (2019) 941–964.
[7] X. Fu, W.K. Ma, J.M. Bioucas-Dias, T.H. Chan, Semiblind hyperspectral unmixing in the presence of spectral library mismatches, IEEE Trans. Geosci. Remote Sens. 54 (9) (2016) 5171–5184.
[8] S. Gandy, B. Recht, I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, Inverse Prob. 27 (2) (2011) 025010.
[9] D. Goldfarb, Z.T. Qin, Robust low-rank tensor recovery: models and algorithms, SIAM J. Matrix Anal. Appl. 35 (1) (2014) 225–253.
[10] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, Non-local meets global: an integrated paradigm for hyperspectral denoising, in: In the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 6868–6877.
[11] C.J. Hillar, L.-H. Lim, Most tensor problems are NP-hard, J. ACM 60 (6) (2013), 45:1–45:39.
[12] W. Hu, D. Tao, W. Zhang, Y. Xie, Y. Yang, The twist tensor nuclear norm for video completion, IEEE Trans. Neural Netw. Learn. Syst. 28 (12) (2017) 2961–2973.
[13] Y.-M. Huang, H.-Y. Yan, Y.-W. Wen, X. Yang, Rank minimization with applications to image noise removal, Inf. Sci. 429 (2018) 147–163.

[8] http://lesun.weebly.com/hyperspectral-data-set.html.

[14] T.-Y. Ji, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, L.-J. Deng, A non-convex tensor rank approximation for tensor completion, Appl. Math. Model. 48 (2017) 410–422.
[15] T.-Y. Ji, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, G. Liu, Tensor completion using total variation and low-rank matrix factorization, Inf. Sci. 326 (2016) 243–257.
[16] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm, J. Comput. Appl. Math. (2019), https://doi.org/10.1016/j.cam.2019.112680.
[17] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, T.-Y. Ji, L.-J. Deng, Matrix factorization for low-rank tensor completion using framelet prior, Inf. Sci. 436 (2018) 403–417.
[18] I. Kajo, N. Kamel, Y. Ruichek, A.S. Malik, SVD-based tensor-completion technique for background initialization, IEEE Trans. Image Process. 27 (6) (2018) 3114–3126.
[19] M.E. Kilmer, K. Braman, N. Hao, R.C. Hoover, Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging, SIAM J. Matrix Anal. Appl. 34 (1) (2013) 148–172.
[20] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[21] M. Li, Q. Xie, Q. Zhao, W. Wei, S. Gu, J. Tao, D. Meng, Video rain streak removal by multiscale convolutional sparse coding, in: In the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6644–6653.
[22] S. Li, R. Dian, L. Fang, J.M. Bioucas-Dias, Fusing hyperspectral and multispectral images via coupled sparse tensor factorization, IEEE Trans. Image Process. 27 (8) (2018) 4118–4130.
[23] X. Li, Y. Ye, X. Xu, Low-rank tensor completion with total variation for visual data in painting, in: AAAI, 2017, pp. 2210–2216..
[24] J. Liu, P. Musialski, P. Wonka, J. Ye, Tensor completion for estimating missing values in visual data, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 208–220.
[25] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, IEEE Trans. Pattern Anal. Mach. Intell. (2019), https://doi.org/10.1109/TPAMI.2019.2891760.
[26] B. Madathil, S.N. George, Twist tensor total variation regularized-reweighted nuclear norm based tensor completion for video missing area recovery, Inf. Sci. 423 (2018) 376–397.
[27] A. Mian, R. Hartley, Hyperspectral video restoration using optical flow and sparse coding, Opt. Express 20 (10) (2012) 10658–10673.
[28] C. Mu, B. Huang, J. Wright, D. Goldfarb, Square deal: Lower bounds and improved relaxations for tensor recovery, in: In the International Conference on Machine Learning (ICML), 2014, pp. 73–81.
[29] M.K. Ng, Q. Yuan, L. Yan, J. Sun, An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data, IEEE Trans. Geosci. Remote Sens. 55 (6) (2017) 3367–3381.
[30] Y. Peng, J. Suo, Q. Dai, W. Xu, Reweighted low-rank matrix recovery and its application in image restoration, IEEE Trans. Cybern. 44 (12) (2014) 2418–2430.
[31] O. Semerci, N. Hao, M.E. Kilmer, E.L. Miller, Tensor-based formulation and nuclear norm regularization for multienergy computed tomography, IEEE Trans. Image Process. 23 (4) (2014) 1678–1693.
[32] Y. Wang, D. Meng, M. Yuan, Sparse recovery: from vectors to tensors, Natl. Sci. Rev. 5 (5) (2018) 756–767.
[33] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[34] W. Wei, L. Yi, Q. Xie, Q. Zhao, D. Meng, Z. Xu, Should we encode rain streaks in video as deterministic or stochastic?, in: In the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp 2516–2525.
[35] Q. Xie, Q. Zhao, D. Meng, Z. Xu, Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery, IEEE Trans. Pattern Anal. Mach. Intell. 40 (8) (2018) 1888–1902.
[36] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, Y. Qu, On unifying multi-view self-representations for clustering by tensor multi-rank minimization, Int. J. Comput. Vis. (2016) 1–23..
[37] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, T.-Z. Huang, Low-rank tensor train for tensor robust principal component analysis, Appl. Math. Comput. 367 (2020) 124783.
[38] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, M. Ding, Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization, J. Comput. Appl. Math. 363 (2020) 124–144.
[39] Q. Yao, J.T.-Y. Kwok, B. Han, Efficient nonconvex regularized tensor completion with structure-aware proximal iterations, in: In the International Conference on Machine Learning (ICML), vol. 97, 2019, pp. 7035–7044..
[40] Q. Yao, J.T.-Y. Kwok, T. Wang, T. Liu, Large-scale low-rank matrix learning with nonconvex regularizers, IEEE Trans. Pattern Anal. Mach. Intell. 41 (11) (2019) 2628–2643.
[41] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[42] X. Zhang, A nonconvex relaxation approach to low-rank tensor completion, IEEE Trans. Neural Netw. Learn. Syst. 30 (6) (2019) 1659–1671.
[43] Z. Zhang, S. Aeron, Exact tensor completion using t-SVD, IEEE Trans. Signal Process. 65 (6) (2017) 1511–1526.
[44] Q. Zhao, L. Zhang, A. Cichocki, Bayesian CP factorization of incomplete tensors with automatic rank determination, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1751–1763.
[45] X.-L. Zhao, W.-H. Xu, T.-X. Jiang, Y. Wang, M.K. Ng, Deep plug-and-play prior for low-rank tensor completion, Neurocomputing, to be published. https://doi.org/10.1016/j.neucom.2020.03.018..
[46] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Y. Chen, W. He, Double factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image, IEEE Trans. Geosci. Remote Sens., to be published.https://doi.org/10.1109/TGRS.2020.2987954..
[47] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, T.-Y. Ji, Mixed noise removal in hyperspectral image via low-fibered-rank regularization, IEEE Trans. Geosci. Remote Sens. 58 (1) (2020) 734–749.
[48] P. Zhou, C. Lu, Z. Lin, C. Zhang, Tensor factorization for low-rank tensor completion, IEEE Trans. Image Process. 27 (3) (2018) 1152–1163.
[49] L. Zhuang, J.M. Bioucas-Dias, Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 11 (3) (2018) 730–742.