

# Mixed Noise Removal in Hyperspectral Image via Low-Fibered-Rank Regularization

Yu-Bang Zheng<sup>ID</sup>, Ting-Zhu Huang<sup>ID</sup>, Xi-Le Zhao<sup>ID</sup>, Tai-Xiang Jiang<sup>ID</sup>, Tian-Hui Ma<sup>ID</sup>, and Teng-Yu Ji<sup>ID</sup>

**Abstract**—The tensor tubal rank, defined based on the tensor singular value decomposition (t-SVD), has obtained promising results in hyperspectral image (HSI) denoising. However, the framework of the t-SVD lacks flexibility for handling different correlations along different modes of HSIs, leading to suboptimal denoising performance. This article mainly makes three contributions. First, we introduce a new tensor rank named tensor fibered rank by generalizing the t-SVD to the mode- $k$  t-SVD, to achieve a more flexible and accurate HSI characterization. Since directly minimizing the fibered rank is NP-hard, we suggest a three-directional tensor nuclear norm (3DTNN) and a three-directional log-based tensor nuclear norm (3DLogTNN) as its convex and nonconvex relaxation to provide an efficient numerical solution, respectively. Second, we propose a fibered rank minimization model for HSI mixed noise removal, in which the underlying HSI is modeled as a low-fibered-rank component. Third, we develop an efficient alternating direction method of multipliers (ADMMs)-based algorithm to solve the proposed model, especially, each subproblem within ADMM is proven to have a closed-form solution, although 3DLogTNN is nonconvex. Extensive experimental results demonstrate that the proposed method has superior denoising performance, as compared with the state-of-the-art competing methods on low-rank matrix/tensor approximation and noise modeling.

**Index Terms**—Alternating direction method of multipliers (ADMMs), hyperspectral image (HSI), log-based function, tensor fibered rank, tensor nuclear norm.

## I. INTRODUCTION

**H**YPERSPECTRAL images (HSIs) provide wealthy spatial and spectral information of real scenes and have been widely used in many applications [1], [2], such as mineral

detection, earth observation, and environmental monitoring. However, due to imaging limitations, HSIs inevitably suffer from various noises, such as Gaussian noise, salt and pepper noise, and stripe noise. HSI denoising aims at estimating the clean HSI from its degraded observation, which is a critical preprocessing step for many subsequent applications, such as target detection [3]–[6], classification [7]–[9], and unmixing [10]–[13]. The central issue of HSI denoising is to exploit the spatial and spectral prior knowledge of HSIs. Generally, a typical HSI priors include piecewise smoothness, nonlocal self-similarity, and low rankness [14]–[26].

The piecewise smoothness focuses on the local continuity of HSIs along both the spatial and spectral modes. To exploit this prior, many HSI denoising models are built on the total variation (TV) regularization, such as spatial–spectral TV (SSTV) [27], spectral–spatial adaptive TV [28], tensor decomposition-based SSTV [29], and enhanced 3-D TV [30]. Not limited to the above-mentioned methods that constrain the HSI directly, Chen *et al.* [31] incorporated the spatial piecewise smoothness prior of HSIs into a low-rank matrix factorization (LRMF) framework, where one of the factor matrices is characterized by framelet-based regularization.

The nonlocal self-similarity means that HSIs contain many similar 2-D patches/3-D cubes at different locations. Early methods treat each band of HSIs as a gray-level image and impose the spatial nonlocal self-similar prior to remove the noise band-by-band, such as the nonlocal means filter-based method [32], dictionary learning-based method [33], BM3D [34], and weighted nuclear norm minimization-based method [35]. However, this practice ignores the correlations between different spectral bands, and thus, usually cannot exhaust the estimation potential of the nonlocal self-similar prior. To tackle this issue, many denoising methods considered the spatial–spectral nonlocal self-similar prior of HSIs, i.e., consider similar 3-D cubes, rather than 2-D patches, as the basic unit of denoising, such as BM4D [36], MSPCA-BM3D [37], nonlocal tensor dictionary learning-based method [38], and intrinsic tensor sparsity regularization-based method [39].

The low rankness reflects the global correlation in the spectral–spatial domain of HSIs. The main idea of low-matrix-rank-based methods [40]–[48] is that unfold the target HSI as a matrix by vectorizing each band as a column and then promote the low rankness of the unfolding matrix. However, the unfolding operator destroys the spatial information of HSIs. Recently, many denoising methods consider the tensor low rankness of

Manuscript received April 29, 2019; revised July 10, 2019 and August 11, 2019; accepted September 6, 2019. Date of publication October 1, 2019; date of current version December 27, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61772003, Grant 61876203, and Grant 11901450, in part by the Fundamental Research Funds for the Central Universities under Grant 31020180QD126, in part by the National Postdoctoral Program for Innovative Talents under Grant BX20180252, and in part by the Project funded by China Postdoctoral Science Foundation under Grant 2018M643611. (Corresponding authors: Ting-Zhu Huang; Xi-Le Zhao.)

Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, and T.-X. Jiang are with the Research Center for Image and Vision Computing, School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhengyubang@163.com; tingzhuhuang@126.com; xilzhao122003@163.com; taixiangjiang@gmail.com).

T.-H. Ma is with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: nkmth0307@126.com).

T.-Y. Ji is with the School of Science, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: tengyu\_j66@126.com).

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

Digital Object Identifier 10.1109/TGRS.2019.2940534

HSIs, based on tensor decompositions and the corresponding tensor rank. Two typical tensor decompositions are the Tucker decomposition and PARAFAC decomposition. Representative methods based on Tucker decomposition include the low-rank tensor approximation (LRTA) [49], the multidimensional filtering [50], the Tucker rank and TV minimization [51], and the nonlocal-based low-Tucker-rank approximation [52]. Representative methods based on the PARAFAC decomposition include the PARAFAC decomposition [53] and the rank-1 tensor decomposition [54].

To highlight the contributions and the innovations of the proposed method, we give a detailed review of related works on an HSI-mixed noise removal based on low-rank modeling.

### A. Related Work

Based on the success of low-rank matrix recovery (LRMR), Zhang *et al.* [43] developed a mixed noise removal method for HSI. This method first divides the HSI into full-band 3-D cubes, then unfolds each 3-D cube to a matrix by vectorizing each band as a column, and finally removes the mixed noise by promoting the low rankness of the unfolding matrix and the sparsity of the non-Gaussian noise. To better promote the low rankness, Xie *et al.* [44] applied the weighted Schatten  $p$ -norm to relax the rank of the unfolding matrix. Since different bands are usually degraded by noises with different intensity in real HSIs, He *et al.* [45] proposed a patchwise low-rank matrix approximation-based method, which designs a noise-adjusted iteration framework based on the noise intensity of each band to remove the mixed noises. Not limited to exploring the priors of the target HSIs, several works utilize flexible distributions to model the noise. For example, under the framework of LRMF, Cao *et al.* [47] employed a mixture of independent and identically distributed (i.i.d.) exponential power distributions assumption and Chen *et al.* [55] utilized non-i.i.d. mixture of Gaussian noise assumption to model the noises, respectively. However, the unfolding operator, involved in the above-mentioned matrix-based methods, will destroy the spatial information of HSIs, thus making these methods difficult to preserve the intrinsic structure of the target HSI. As HSIs can be regarded as three-way tensors, the tensor low-rankness characterization for HSIs is expected to explore the global correlation and preserve the intrinsic structure information.

The recent tensor tubal rank, defined based on the tensor singular value decomposition (t-SVD), obtains promising results in LRTA, such as tensor completion [56], [57] and tensor robust principal component analysis (TRPCA) [58], [59]. TRPCA has shown its effectiveness to remove the sparse noise, while it is incapable of removing the Gaussian noise. To tackle the mixed noise in HSIs, Fan *et al.* [60] proposed an HSI mixed noise removal model based on low-tubal-rank tensor recovery (LRTR), which decomposes the noisy HSI into three parts, i.e., a low-tubal-rank part (the clean HSI), a Gaussian noise part, and a sparse noise part. However, the t-SVD and the induced tubal rank lack of flexibility for handling different correlations along different modes of HSIs. For instance, when setting the band of an HSI to be the frontal

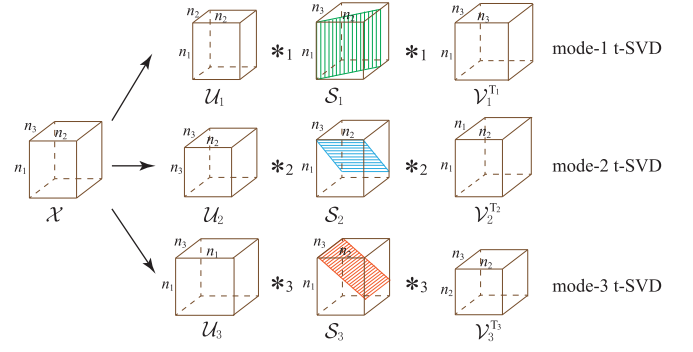


Fig. 1. Illustration of the mode- $k$  t-SVD of  $\mathcal{X}$  for  $k = 1, 2, 3$ .

slice of a three-way tensor, the t-SVD characterizes its spatial correlations via SVDs, while describes its spectral correlation by the embedded circular convolution. This inflexible HSI characterization usually results in suboptimal denoising performance. Naturally, treating each mode flexibly is expected to remedy this defect.

### B. Contributions

This article mainly has three contributions.

First, we propose a novel tensor decomposition by generalizing the t-SVD to the mode- $k$  t-SVD, which achieves a more flexible and accurate HSI low-rankness characterization. Mathematically, the mode- $k$  t-SVD factorizes a three-way tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  tensors as

$$\mathcal{X} = \mathcal{U}_k *_{\mathcal{S}_k} \mathcal{V}_k^{\mathcal{T}_k}, \quad k = 1, 2, 3$$

where  $*_{\mathcal{S}_k}$  is the mode- $k$  t-product,  $\mathcal{U}_k$  and  $\mathcal{V}_k$  are the mode- $k$  orthogonal tensors,  $(\cdot)^{\mathcal{T}_k}$  is the mode- $k$  conjugate transposition, and  $\mathcal{S}_k \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is the mode- $k$  diagonal tensor, the tensor whose mode- $k$  slices are diagonal matrices (see details in Section III-A); see Fig. 1 for an illustration for the mode- $k$  t-SVD scheme. Based on this novel decomposition, we propose a new tensor rank, termed tensor fibered rank, which is defined as a vector whose  $k$ th element is the number of nonzero mode- $k$  fibers of  $\mathcal{S}_k$ . Taking the HSI *Washington DC Mall*, shown in Fig. 2 as an example, we observe that slices along all modes are approximately low rank, which suggests that both spatial and spectral correlations can be evidently observed both quantitatively and visually. And such correlations can be exactly shown by the proposed fibered rank.

Second, we incorporate the proposed low-fibered-rank prior into a novel HSI denoising model, which can estimate the underlying HSI from its observation degraded by a mix of various noises. Mathematically, the proposed model is formulated as

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{N}, \mathcal{S}} \quad & \text{rank}_{\mathcal{F}}(\mathcal{X}) + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{S}\|_1 \\ \text{s.t.} \quad & \mathcal{Y} = \mathcal{X} + \mathcal{N} + \mathcal{S} \end{aligned}$$

where  $\mathcal{X}$  is the underlying HSI,  $\mathcal{Y}$  is the observed HSI,  $\mathcal{N}$  is the Gaussian noise,  $\mathcal{S}$  is the sparse noise, and  $\lambda_1$  and  $\lambda_2$  are regularization parameters. Since directly minimizing

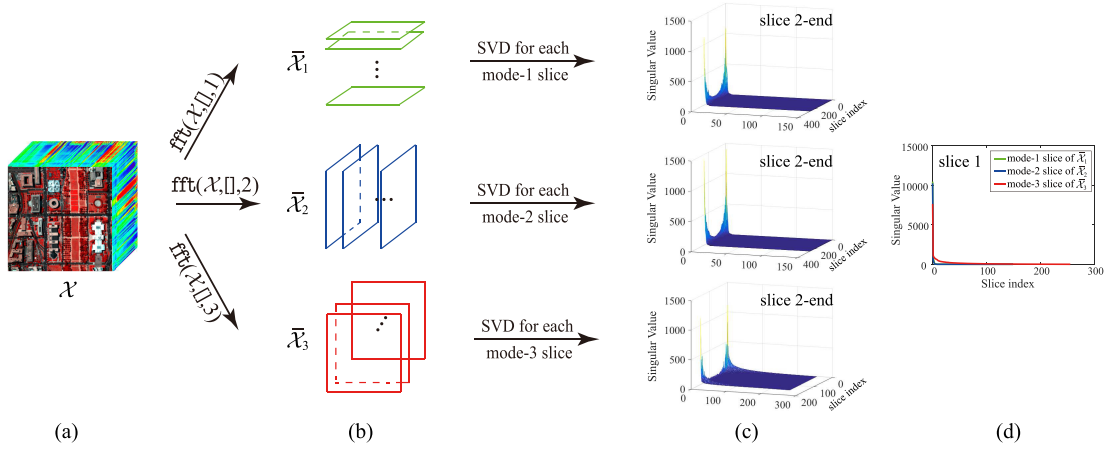


Fig. 2. Illustration of low-rank (tensor fibered rank) priors of an HSI. (a) HSI *Washington DC Mall* of size  $256 \times 256 \times 150$ . (b) Generated tensor  $\tilde{\mathcal{X}}_k$  by performing the DFT along each mode- $k$  fiber of  $\mathcal{X}$ . (c) Singular value curves from the second to the end mode- $k$  slices of  $\tilde{\mathcal{X}}_k$ . (d) Singular value curves of the first mode- $k$  slices of  $\tilde{\mathcal{X}}_k$ .

the proposed fibered rank is NP-hard, we introduce a three-directional tensor nuclear norm (3DTNN) as its convex relaxation to provide an efficient numerical solution. In addition, we also suggest a three-directional log-based tensor nuclear norm (3DLogTNN) as its nonconvex relaxation to better promote the low rankness of the solution.

Third, we design an efficient alternating direction method of multipliers (ADMMs)-based algorithm [61], [62] to solve the proposed 3DTNN and 3DLogTNN-based models. Extensive experiments carried out on simulated and real-world noisy HSIs demonstrate the efficiency of the proposed algorithm and the superiority of the proposed method over the related state-of-the-art ones.

The significant feature of the proposed method is the flexible and accurate HSI low-rankness characterization. Since the unfolding operation is not involved, the proposed method can better preserve the intrinsic structure than the low-matrix-rank-based methods. In addition, the proposed method achieves a better denoising performance than the low-tubal-rank-based methods, since it can directly and fully exploit both the spatial and the spectral correlations.

This article is an extension of the material published in [63]. The new materials are the following: 1) theoretical results of the tensor fibered rank and its related concepts are provided for better understanding, and a new nonconvex relaxation 3DLogTNN is introduced to show the tensor fibered rank more accurately and treat singular values differently; 2) a 3DLogTNN-based HSI mixed noise removal model is proposed with an ADMM-based solver, especially each subproblem within ADMM is proven to have a closed-form solution; and 3) extensive experimental results demonstrate that the 3DLogTNN-based model significantly outperforms the compared ones, including the 3DTNN-based model.

### C. Organization

The outline of this article is as follows. Section II presents some notations. Section III generalizes some definitions, designs the mode- $k$  t-SVD, and defines the fibered rank.

Section IV proposes the 3DTNN-based HSI denoising model and the 3DLogTNN-based HSI denoising model with an ADMM-based solver. Section V evaluates the performance of the proposed method. Section VI concludes this article.

## II. NOTATIONS

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.,  $\mathcal{X}$ ). For a three-way tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , with the MATLAB notation, we denote its  $(i, j, s)$ th element as  $\mathcal{X}(i, j, s)$ , 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. For the sake of clarity, we use  $X_1^{(i)} \in \mathbb{R}^{n_2 \times n_3}$ ,  $X_2^{(i)} \in \mathbb{R}^{n_3 \times n_1}$ , and  $X_3^{(i)} \in \mathbb{R}^{n_1 \times n_2}$  to denote the  $i$ th mode-1 (horizontal), mode-2 (lateral), and mode-3 (frontal) slices of  $\mathcal{X}$ , respectively. The Frobenius norm of  $\mathcal{X}$  is defined as  $\|\mathcal{X}\|_F := (\sum_{i,j,s} |\mathcal{X}(i, j, s)|^2)^{1/2}$ . The  $\ell_1$  norm of  $\mathcal{X}$  is defined as  $\|\mathcal{X}\|_1 := (\sum_{i,j,s} |\mathcal{X}(i, j, s)|)$ . We use  $\tilde{\mathcal{X}}_k$  to denote the tensor generated by performing Discrete Fourier Transformation (DFT) along each mode- $k$  fibers of  $\mathcal{X}$ , i.e.,  $\tilde{\mathcal{X}}_k = \text{fft}(\mathcal{X}, [], k)$ . Naturally, we can compute  $\mathcal{X}$  via  $\mathcal{X} = \text{ifft}(\tilde{\mathcal{X}}_k, [], k)$ .

The mode- $k$  unfolding of an  $N$ -way tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_N}$  is denoted as  $X_{(k)} \in \mathbb{R}^{n_k \times \prod_{i \neq k} n_i}$ , whose  $(i_k, j)$ th element maps to the  $(i_1, i_2, \dots, i_N)$ th element of  $\mathcal{X}$ , where  $j = 1 + \sum_{s=1, s \neq k}^N (i_s - 1)J_s$  with  $J_s = \sum_{m=1, m \neq k}^{s-1} n_m$ . The unfolding operator and its inversion are denoted as “unfold” and “fold,” respectively, i.e.,  $X_{(k)} = \text{unfold}(\mathcal{X}, k)$  and  $\mathcal{X} = \text{fold}(X_{(k)}, k)$ .

## III. MODE- $k$ t-SVD AND FIBERED RANK

In this section, we generalize related definitions, design the mode- $k$  t-SVD, and define tensor fibered rank.

### A. Generalized Definitions

To introduce our results, we generalize five block-based operations and some definitions [56] related to this article. Let  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  be a three-way tensor, and the mode- $k$

block circulation operation is defined as

$$\text{bcirc}(\mathcal{X}, k) := \begin{pmatrix} X_k^{(1)} & X_k^{(n_k)} & \dots & X_k^{(2)} \\ X_k^{(2)} & X_k^{(1)} & \dots & X_k^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ X_k^{(n_k)} & X_k^{(n_k-1)} & \dots & X_k^{(1)} \end{pmatrix}$$

where  $X_k^{(i)}$  denotes the  $i$ th mode- $k$  slice of  $\mathcal{X}$ .

The mode- $k$  block diagonalization operation and its inverse operation are defined as

$$\text{bdiag}(\mathcal{X}, k) := \begin{pmatrix} X_k^{(1)} & & & \\ & X_k^{(2)} & & \\ & & \ddots & \\ & & & X_k^{(n_k)} \end{pmatrix}$$

$$\text{bdfold}(\text{bdiag}(\mathcal{X}, k), k) := \mathcal{X}.$$

The mode- $k$  block vectorization operation and its inverse operation are defined as

$$\text{bvec}(\mathcal{X}, k) := \begin{pmatrix} X_k^{(1)} \\ X_k^{(2)} \\ \vdots \\ X_k^{(n_k)} \end{pmatrix}, \quad \text{bvfold}(\text{bvec}(\mathcal{X}, k), k) := \mathcal{X}.$$

**Definition 1 (Mode- $k$  t-Product):** The mode-1 t-product between  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and  $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_3 \times n_4}$  is defined as

$$\mathcal{X} *_1 \mathcal{Y} := \text{bvfold}(\text{bcirc}(\mathcal{X}, 1) \text{bvec}(\mathcal{Y}, 1), 1) \in \mathbb{R}^{n_1 \times n_2 \times n_4}.$$

The mode-2 t-product between  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and  $\mathcal{Y} \in \mathbb{R}^{n_4 \times n_2 \times n_1}$  is defined as

$$\mathcal{X} *_2 \mathcal{Y} := \text{bvfold}(\text{bcirc}(\mathcal{X}, 2) \text{bvec}(\mathcal{Y}, 2), 2) \in \mathbb{R}^{n_4 \times n_2 \times n_3}.$$

The mode-3 t-product between  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and  $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$  is defined as

$$\mathcal{X} *_3 \mathcal{Y} := \text{bvfold}(\text{bcirc}(\mathcal{X}, 3) \text{bvec}(\mathcal{Y}, 3), 3) \in \mathbb{R}^{n_1 \times n_4 \times n_3}.$$

Indeed, the mode- $k$  t-product can be regarded as a matrix-matrix multiplication, except that the multiplication operation between scalars is replaced by circular convolution between the mode- $k$  fibers, that is,

$$\mathcal{F} = \mathcal{X} *_1 \mathcal{Y} \Leftrightarrow \mathcal{F}(:, j, s) = \sum_{t=1}^{n_3} \mathcal{X}(:, j, t) \star \mathcal{Y}(:, t, s)$$

$$\mathcal{F} = \mathcal{X} *_2 \mathcal{Y} \Leftrightarrow \mathcal{F}(i, :, s) = \sum_{t=1}^{n_1} \mathcal{X}(t, :, s) \star \mathcal{Y}(i, :, t)$$

$$\mathcal{F} = \mathcal{X} *_3 \mathcal{Y} \Leftrightarrow \mathcal{F}(i, j, :) = \sum_{t=1}^{n_2} \mathcal{X}(i, t, :) \star \mathcal{Y}(t, j, :)$$

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

$$\tilde{\mathcal{F}}_k = \text{bdfold}(\text{bdiag}(\tilde{\mathcal{X}}_k, k) \text{bdiag}(\tilde{\mathcal{Y}}_k, k), k).$$

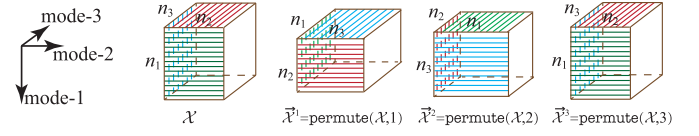


Fig. 3. Illustration of the mode- $k$  permutation of an  $n_1 \times n_2 \times n_3$  tensor.

**Definition 2 (Mode- $k$  Conjugate Transpose):** Assuming that  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is a three-way tensor. the mode- $k$  conjugate transpose of it, denote as  $\mathcal{X}^{\text{T}_k}$ , is the tensor obtained by conjugate transposing each of the mode- $k$  slices and then reversing the order of transposed mode- $k$  slices 2 through  $n_k$ .

**Definition 3 (Mode- $k$  Identity Tensor):** The mode- $k$  identity tensor  $\mathcal{I}_k \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is the tensor whose first mode- $k$  slice is an identity matrix and other mode- $k$  slices are all zeros.

**Definition 4: [Mode- $k$  Orthogonal Tensor]** A three-way tensor  $\mathcal{Q}$  is mode- $k$  orthogonal if

$$\mathcal{Q} *_k \mathcal{Q}^{\text{T}_k} = \mathcal{Q}^{\text{T}_k} *_k \mathcal{Q} = \mathcal{I}_k.$$

**Definition 5 (Mode- $k$  Diagonal Tensor):** A three-way tensor  $\mathcal{S}$  is mode- $k$  diagonal if each of its mode- $k$  slices is a diagonal matrix.

**Definition 6 (Tensor Mode- $k$  Permutation):** Assuming that  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is a three-way tensor. The mode- $k$  permutation of  $\mathcal{X}$ , denoted as  $\tilde{\mathcal{X}}^k$ , is defined as the tensor whose  $i$ th mode-3 slice is the  $i$ th mode- $k$  slice of  $\mathcal{X}$ , i.e.,  $\mathcal{X}(i, j, s) = \tilde{\mathcal{X}}^1(j, s, i) = \tilde{\mathcal{X}}^2(s, i, j) = \tilde{\mathcal{X}}^3(i, j, s)$ . We define the corresponding operation as  $\tilde{\mathcal{X}}^k := \text{permute}(\mathcal{X}, k)$  and its inverse operation as  $\mathcal{X} := \text{ipermute}(\tilde{\mathcal{X}}^k, k)$ .

Fig. 3 shows the mode- $k$  permutation of an  $n_1 \times n_2 \times n_3$  tensor. With the above-mentioned definitions, we introduce Theorem 1 for the design of the mode- $k$  t-SVD.

**Theorem 1:** Assuming that  $\mathcal{X}$  and  $\mathcal{Y}$  are two three-way tensors. Then, the properties hold as follows.

- 1)  $\text{bcirc}(\tilde{\mathcal{X}}^k, 3) = \text{bcirc}(\mathcal{X}, k)$ .
- 2)  $\text{bvec}(\tilde{\mathcal{X}}^k, 3) = \text{bvec}(\mathcal{X}, k)$ .
- 3)  $\tilde{\mathcal{X}}^k = \text{bvfold}(\text{bvec}(\mathcal{X}, k), 3)$ .
- 4)  $(\mathcal{X}^{\text{T}_k})^k = (\tilde{\mathcal{X}}^k)^{\text{T}_3}$ .
- 5)  $\mathcal{X}$  is the mode- $k$  identity tensor if and only if  $\tilde{\mathcal{X}}^k$  is the mode-3 identity tensor.
- 6)  $\mathcal{X}$  is the mode- $k$  diagonal tensor if and only if  $\tilde{\mathcal{X}}^k$  is the mode-3 diagonal tensor.
- 7)  $\mathcal{F} = \mathcal{X} *_k \mathcal{Y}$  if and only if  $\tilde{\mathcal{F}}^k = \tilde{\mathcal{X}}^k *_3 \tilde{\mathcal{Y}}^k$ .
- 8)  $\mathcal{X}$  is the mode- $k$  orthogonal tensor if and only if  $\tilde{\mathcal{X}}^k$  is the mode-3 orthogonal tensor.

**Proof:** The first six properties can be directly derived by the above-mentioned definitions. Now, we prove the last two properties.

For property (7), let  $\mathcal{F} = \mathcal{X} *_k \mathcal{Y}$ , that is

$$\mathcal{F} = \text{bvfold}(\text{bcirc}(\mathcal{X}, k) \text{bvec}(\mathcal{Y}, k), k)$$

$$= \text{bvfold}(\text{bcirc}(\tilde{\mathcal{X}}^k, 3) \text{bvec}(\tilde{\mathcal{Y}}^k, 3), k).$$



**Algorithm 1** Mode- $k$  t-SVD for Three-Way Tensors**Input:**  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and  $k$ .

- 1:  $\tilde{\mathcal{X}} \leftarrow \text{fft}(\mathcal{X}, [], k)$ .
- 2: **for**  $i = 1$  to  $n_k$  **do**
- 3:  $[U, S, V] = \text{svd}((\tilde{\mathcal{X}}_k^{(i)})_k)$ .
- 4:  $(\tilde{\mathcal{U}}_k)_k^{(i)} \leftarrow U$ ;  $(\tilde{\mathcal{S}}_k)_k^{(i)} \leftarrow S$ ;  $(\tilde{\mathcal{V}}_k)_k^{(i)} \leftarrow V$ .
- 5: **end for**
- 6:  $\mathcal{U}_k \leftarrow \text{ifft}(\tilde{\mathcal{U}}_k, [], k)$ .
- 7:  $\mathcal{S}_k \leftarrow \text{ifft}(\tilde{\mathcal{S}}_k, [], k)$ .
- 8:  $\mathcal{V}_k \leftarrow \text{ifft}(\tilde{\mathcal{V}}_k, [], k)$ .

**Output:**  $\mathcal{U}_k, \mathcal{S}_k, \mathcal{V}_k$ .

Thus,  $\text{bvec}(\mathcal{F}, k) = \text{bcirc}(\tilde{\mathcal{X}}^k, 3)\text{bvec}(\tilde{\mathcal{Y}}^k, 3)$ , and with property (4), we can obtain

$$\begin{aligned} \tilde{\mathcal{F}}^k &= \text{bvfold}(\text{bvec}(\mathcal{F}, k), 3) \\ &= \text{bvfold}(\text{bcirc}(\tilde{\mathcal{X}}^k, 3)\text{bvec}(\tilde{\mathcal{Y}}^k, 3), 3) \\ &= \tilde{\mathcal{X}}^k *_3 \tilde{\mathcal{Y}}^k. \end{aligned}$$

Conversely, if  $\tilde{\mathcal{F}}^k = \tilde{\mathcal{X}}^k *_3 \tilde{\mathcal{Y}}^k$ , we can obtain  $\mathcal{F} = \mathcal{X} *_k \mathcal{Y}$ .

For property (8), if  $\mathcal{X}$  is the mode- $k$  orthogonal tensor, that is,

$$\mathcal{X} *_k \mathcal{X}^{\text{T}_k} = \mathcal{X}^{\text{T}_k} *_k \mathcal{X} = \mathcal{I}_k$$

with property (7), we can obtain

$$\tilde{\mathcal{X}}^k *_3 (\tilde{\mathcal{X}}^{\text{T}_k})^k = (\tilde{\mathcal{X}}^{\text{T}_k})^k *_3 \tilde{\mathcal{X}}^k = (\mathcal{I}_k)^k$$

which can be rewritten as

$$\tilde{\mathcal{X}}^k *_3 (\tilde{\mathcal{X}}^k)^{\text{T}_3} = (\tilde{\mathcal{X}}^k)^{\text{T}_3} *_3 \tilde{\mathcal{X}}^k = (\mathcal{I}_k)^k.$$

Note that  $\mathcal{I}_k$  is the mode- $k$  identity tensor, so  $(\mathcal{I}_k)^k$  is the mode-3 identity tensor. Thus,  $\tilde{\mathcal{X}}^k$  is a mode-3 orthogonal tensor. Conversely, if  $\tilde{\mathcal{X}}^k$  is the mode-3 orthogonal tensor, we can obtain that  $\mathcal{X}$  is a mode- $k$  orthogonal tensor.  $\square$

**B. Mode- $k$  t-SVD and Fibered Rank**

As mentioned in Section I, the t-SVD lacks the flexibility to handle different correlations along different modes of HSIs, leading to inadequate characterization for the spectral rank. To tackle this defect, we design the following mode- $k$  t-SVD.

**Theorem 2 (Mode- $k$  t-SVD):** A three-way tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  has the factorization

$$\mathcal{X} = \mathcal{U}_k *_k \mathcal{S}_k *_k \mathcal{V}_k^{\text{T}_k}, \quad k = 1, 2, 3$$

where  $\mathcal{U}_k$  and  $\mathcal{V}_k$  are the mode- $k$  orthogonal tensors and  $\mathcal{S}_k \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is the mode- $k$  diagonal tensor.

*Proof.* When  $k = 3$ , the mode-3 t-SVD is actually the t-SVD, according to [58, Definition 2.7],  $\mathcal{X}$  can be factored as  $\mathcal{X} = \mathcal{U}_3 *_3 \mathcal{S}_3 *_3 \mathcal{V}_3^{\text{T}_3}$ . Now, we prove the cases of  $k = 1$  and  $k = 2$ . For the tensor  $\tilde{\mathcal{X}}^k$  ( $k = 1, 2$ ), it can be factored as

$$\tilde{\mathcal{X}}^k = (\tilde{\mathcal{U}}_k)^k *_3 (\tilde{\mathcal{S}}_k)^k *_3 (\tilde{\mathcal{V}}_k)^k.$$

With property (4) in Theorem 1, we obtain that

$$\tilde{\mathcal{X}}^k = (\tilde{\mathcal{U}}_k)^k *_3 (\tilde{\mathcal{S}}_k)^k *_3 (\tilde{\mathcal{V}}_k)^{\text{T}_k}.$$

TABLE I  
RANK ESTIMATION OF TWO HSIs

Data	Size	Tucker rank	Fibered rank
Washington DC Mall	$256 \times 256 \times 150$	(107,110,6)	(8,8,182)
Pavia City Center	$200 \times 200 \times 80$	(84,86,4)	(8,8,138)

Let

$$(\overrightarrow{\mathcal{F}}_k)^k = (\overrightarrow{\mathcal{U}}_k)^k *_3 (\overrightarrow{\mathcal{S}}_k)^k$$

then

$$\tilde{\mathcal{X}}^k = (\overrightarrow{\mathcal{F}}_k)^k *_3 ((\mathcal{V}_k)^{\text{T}_k})^k.$$

With property (7) in Theorem 1, we obtain that

$$\mathcal{X} = \mathcal{F}_k *_k \mathcal{V}_k^{\text{T}_k} \text{ and } \mathcal{F}_k = \mathcal{U}_k *_k \mathcal{S}_k.$$

Thus,  $\mathcal{X} = \mathcal{U}_k *_k \mathcal{S}_k *_k \mathcal{V}_k^{\text{T}_k}$  ( $k = 1, 2$ ). Note that  $(\overrightarrow{\mathcal{U}}_k)^k$  and  $(\overrightarrow{\mathcal{V}}_k)^k$  are mode-3 orthogonal tensors and  $(\overrightarrow{\mathcal{S}}_k)^k$  is the mode-3 diagonal tensor, according to properties (6) and (8) in Theorem 1, and  $\mathcal{U}_k$  and  $\mathcal{V}_k$  are mode- $k$  orthogonal tensors and  $\mathcal{S}_k$  is the mode- $k$  diagonal tensor.  $\square$

The mode- $k$  t-SVD can be efficiently obtained by computing a series of matrix SVDs in the Fourier domain, see Algorithm 1. Now, we give the definition of the corresponding tensor mode- $k$  fibered rank and mode- $k$  multirank.

**Definition 7 (Mode- $k$  Tensor Fibered Rank and Multirank):** Let  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  be a three-way tensor. The mode- $k$  fibered rank of  $\mathcal{X}$ , denote as  $\text{rank}_{f_k}(\mathcal{X})$ , is defined as the number of nonzero mode- $k$  fibers of  $\mathcal{S}_k$ , where  $\mathcal{S}_k$  comes from the mode- $k$  t-SVD of  $\mathcal{X}$ :  $\mathcal{X} = \mathcal{U}_k *_k \mathcal{S}_k *_k \mathcal{V}_k^{\text{T}_k}$ . The tensor mode- $k$  multirank of  $\mathcal{X}$  is a vector  $\text{rank}_{m_k}(\mathcal{X}) \in \mathbb{R}^{n_k}$ , whose  $i$ th element is the rank of  $i$ th mode- $k$  slice of  $\tilde{\mathcal{X}}_k$ , where  $\tilde{\mathcal{X}}_k = \text{fft}(\mathcal{X}, [], k)$ . That is,  $\text{rank}_{f_k}(\mathcal{X}) = \max(\text{rank}_{m_k}(\mathcal{X}))$ .

Actually, the tensor tubal rank/multirank is the tensor mode-3 fibered rank/multirank. We define the following tensor fibered rank to combine all mode- $k$  ( $k = 1, 2, 3$ ) fibered rank.

**Definition 8 (Tensor Fibered Rank):** The fibered rank of a three-way tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , denoted as  $\text{rank}_f(\mathcal{X})$ , is defined as a vector, whose  $k$ th element is the mode- $k$  tensor fibered rank.

Table I gives the rank estimation<sup>1</sup> of two HSIs, where the bands of HSIs correspond to the frontal slices of three-way tensors. As observed, the rank of the mode-3 unfolding matrix is much lower than the size of the third mode (spectral mode), which implies a strong correlation along the third mode. According to the tensor fibered rank, such a strong correlation is inadequately shown by the third element (the tubal rank), which can be exactly shown by the other two elements. It demonstrates that compared with the tensor tubal rank, the novel tensor fibered rank achieves a flexible and simultaneous representation for the correlations along different modes.

<sup>1</sup>The rank is approximated by the numbers of the singular values which are larger than 0.01 of the largest one.

#### IV. PROPOSED HSI DENOISING MODELS AND SOLVING ALGORITHM

##### A. Proposed HSI Denoising Models

Considering an observation degraded by a mix of various noises, the proposed HSI denoising model based on low-fibered-rank prior is formulated as

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{N}, \mathcal{S}} \quad & \text{rank}_f(\mathcal{X}) + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{S}\|_1 \\ \text{s.t.} \quad & \mathcal{Y} = \mathcal{X} + \mathcal{N} + \mathcal{S} \end{aligned} \quad (1)$$

where  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is the underlying HSI,  $\mathcal{Y}$  is the observed HSI,  $\mathcal{N}$  is the Gaussian noise,  $\mathcal{S}$  is the sparse noise, and  $\lambda_1$  and  $\lambda_2$  are regularization parameters.

As directly minimizing the tensor fibered rank is NP-hard, we define the following 3DTNN as the convex relaxation of the proposed fibered rank.

**Definition 9 (Mode- $k$  TNN):** The mode- $k$  tensor nuclear norm of a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , denoted as  $\|\mathcal{X}\|_{\text{TNN}_k}$ , is defined as the sum of singular values of all the mode- $k$  slices of  $\bar{\mathcal{X}}_k$ , that is,

$$\|\mathcal{X}\|_{\text{TNN}_k} := \sum_{i=1}^{n_k} \|(\bar{\mathcal{X}}_k)_k^{(i)}\|_*.$$

**Definition 10 (3DTNN):** The 3DTNN of a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , denoted as  $\|\mathcal{X}\|_{3\text{DTNN}}$ , is defined as

$$\|\mathcal{X}\|_{3\text{DTNN}} := \sum_{k=1}^3 \alpha_k \|\mathcal{X}\|_{\text{TNN}_k}$$

where  $\alpha_k \geq 0$  ( $k = 1, 2, 3$ ) and  $\sum_{k=1}^3 \alpha_k = 1$ .

With the mode- $k$  permutational operation, we can obtain  $\|\mathcal{X}\|_{\text{TNN}_k} = \|\bar{\mathcal{X}}^k\|_{\text{TNN}_3}$ . It implies that 3DTNN numerically equals the triple-tubal nuclear norm proposed in [64].

As the convex relaxation of the fibered rank, 3DTNN can provide an efficient numerical solution. However, it has two shortcomings. First, it measures the  $\ell_1$ -norm of nonzero singular values, which is not a good approximation to the fibered rank. Second, it treats each singular value equally, and thus, may not well preserve the major information. This is because the larger singular values usually correspond to the major information, such as contour, sharp edges, and smooth zones, thus should be shrinking less than small singular values [35], [65]–[69]. Therefore, we suggest a 3DLogTNN as a nonconvex relaxation of the proposed fibered rank to overcome the above-mentioned two shortcomings.

**Definition 11 (Mode- $k$  LogTNN):** The mode- $k$  log-based tensor nuclear norm of a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is defined as

$$\text{LogTNN}_k(\mathcal{X}, \varepsilon) := \sum_{i=1}^{n_k} \text{LogMNN}((\bar{\mathcal{X}}_k)_k^{(i)}, \varepsilon)$$

where  $(\bar{\mathcal{X}}_k)_k^{(i)}$  denotes the  $i$ th mode- $k$  slice of  $\bar{\mathcal{X}}_k$  with  $\bar{\mathcal{X}}_k = \text{fft}(\mathcal{X}, [], k)$  and

$$\text{LogMNN}(X, \varepsilon) := \sum_{i=1}^m \log(\sigma_i(X) + \varepsilon).$$

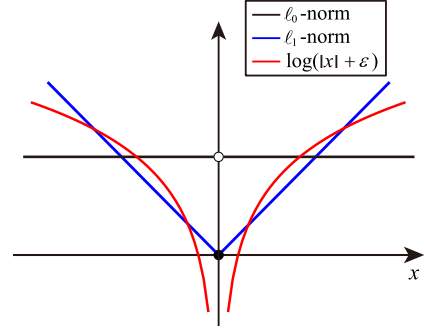


Fig. 4. Comparison of the  $\ell_0$ -norm, the  $\ell_1$ -norm, and the log-function for scalars.

Here  $\sigma_i(X)$  is the  $i$ th singular values of  $X$  and  $\varepsilon > 0$  is a constant.

**Definition 12 (3DLogTNN):** The 3DLogTNN of a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is defined as

$$3\text{DLogTNN}(\mathcal{X}, \varepsilon) := \sum_{k=1}^3 \alpha_k \text{LogTNN}_k(\mathcal{X}, \varepsilon)$$

where  $\alpha_k \geq 0$  ( $k = 1, 2, 3$ ) and  $\sum_{k=1}^3 \alpha_k = 1$ .

Compared to 3DTNN, 3DLogTNN has two advantages. First, it is a closer approximation to the fibered rank than 3DTNN. Fig. 4 shows a comparison of the  $\ell_0$ -norm, the  $\ell_1$ -norm, and the log-function for scalars. We can see that the log-function can better approximate the  $\ell_0$ -norm than the  $\ell_1$ -norm. Thus as the sum of the log-function of singular values, 3DLogTNN can better approximate to the fibered rank than 3DTNN. Second, the tensor singular value thresholding (t-SVT) induced by 3DLogTNN treats singular values differently, by shrinking less the larger ones to keep the major information while shrinking more the smaller ones to suppress noise (detailed theoretical proof will be provided in Section IV-B).

With the definitions of 3DTNN and 3DLogTNN, the proposed 3DTNN-based HSI denoising model and 3DLogTNN-based HSI denoising model are commonly formulated as

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{N}, \mathcal{S}} \quad & \sum_{k=1}^3 \alpha_k \mathcal{F}_k(\mathcal{X}) + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{S}\|_1 \\ \text{s.t.} \quad & \mathcal{Y} = \mathcal{X} + \mathcal{N} + \mathcal{S} \end{aligned} \quad (2)$$

where the function  $\mathcal{F}_k(\mathcal{X})$  are set to be  $\|\mathcal{X}\|_{\text{TNN}_k}$  and  $\text{LogTNN}_k(\mathcal{X}, \varepsilon)$  in 3DTNN-based HSI denoising model and 3DLogTNN-based HSI denoising model, respectively.

##### B. Proposed ADMM-Based Solving Algorithm

We use ADMM to solve (2). We introduce three auxiliary variables  $\mathcal{Z}_k$  ( $k = 1, 2, 3$ ) and reformulate (2) as

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{N}, \mathcal{S}} \quad & \sum_{k=1}^3 \alpha_k \mathcal{F}_k(\mathcal{Z}_k) + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{S}\|_1 \\ \text{s.t.} \quad & \begin{cases} \mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{S}) = 0 \\ \mathcal{X} - \mathcal{Z}_k = 0, \quad k = 1, 2, 3. \end{cases} \end{aligned} \quad (3)$$

The augmented Lagrangian function of (3) is

$$L_{\mu_k, \beta}(\mathcal{Z}_k, \mathcal{X}, \mathcal{N}, \mathcal{S}, \mathcal{M}_k, \mathcal{P}) = \sum_{k=1}^3 \left\{ \alpha_k \mathcal{F}_k(\mathcal{Z}_k) + \langle \mathcal{X} - \mathcal{Z}_k, \mathcal{M}_k \rangle + \frac{\mu_k}{2} \|\mathcal{X} - \mathcal{Z}_k\|_F^2 \right\} + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{S}\|_1 + \langle \mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{S}), \mathcal{P} \rangle + \frac{\beta}{2} \|\mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{S})\|_F^2 \quad (4)$$

where  $\mathcal{M}_k$  ( $k = 1, 2, 3$ ) and  $\mathcal{P}$  are the Lagrange multipliers;  $\mu_k$  ( $k = 1, 2, 3$ ) and  $\beta$  are the penalty parameters. Within the framework of ADMM,  $\mathcal{Z}_k$ ,  $\mathcal{X}$ ,  $\mathcal{N}$ ,  $\mathcal{S}$ ,  $\mathcal{M}_k$ , and  $\mathcal{P}$  are alternately updated as

$$\begin{cases} \mathcal{Z}_k^{p+1} = \arg \min_{\mathcal{Z}_k} L_{\mu_k, \beta}(\mathcal{Z}_k, \mathcal{X}^p, \mathcal{N}^p, \mathcal{S}^p, \mathcal{M}_k^p, \mathcal{P}^p) \\ \mathcal{X}^{p+1} = \arg \min_{\mathcal{X}} L_{\mu_k, \beta}(\mathcal{Z}_k^{p+1}, \mathcal{X}, \mathcal{N}^p, \mathcal{S}^p, \mathcal{M}_k^p, \mathcal{P}^p) \\ \mathcal{N}^{p+1} = \arg \min_{\mathcal{N}} L_{\mu_k, \beta}(\mathcal{Z}_k^{p+1}, \mathcal{X}^{p+1}, \mathcal{N}, \mathcal{S}^p, \mathcal{M}_k^p, \mathcal{P}^p) \\ \mathcal{S}^{p+1} = \arg \min_{\mathcal{S}} L_{\mu_k, \beta}(\mathcal{Z}_k^{p+1}, \mathcal{X}^{p+1}, \mathcal{N}^{p+1}, \mathcal{S}, \mathcal{M}_k^p, \mathcal{P}^p) \\ \mathcal{M}_k^{p+1} = \mathcal{M}_k^p + \mu_k (\mathcal{X}^{p+1} - \mathcal{Z}_k^{p+1}) \\ \mathcal{P}^{p+1} = \mathcal{P}^p + \beta (\mathcal{Y} - \mathcal{X}^{p+1} - \mathcal{N}^{p+1} - \mathcal{S}^{p+1}) \end{cases} \quad (5)$$

The  $\mathcal{Z}_k$  ( $k = 1, 2, 3$ ) subproblems are

$$\mathcal{Z}_k^{p+1} = \arg \min_{\mathcal{Z}_k} \alpha_k \mathcal{F}_k(\mathcal{Z}_k) + \frac{\mu_k}{2} \left\| \mathcal{X}^p - \mathcal{Z}_k + \frac{\mathcal{M}_k^p}{\mu_k} \right\|_F^2. \quad (6)$$

To solve (6), we introduce the following t-SVT.

**Theorem 3 (Mode- $k$  TNN-Based t-SVT):** Assuming that  $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is a three-way tensor, a minimizer to

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

is given by the mode- $k$  TNN-based t-SVT

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

where  $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\text{T}_k}$ ,  $(\bar{\mathcal{S}}_k)^{\tau}_{\text{tnn}} = \max(\bar{\mathcal{S}}_k - \tau, 0)$ ,  $\bar{\mathcal{S}}_k = \text{fft}(\mathcal{S}, [], k)$ , and  $\mathcal{S}_{\text{tnn}}^{\tau} = \text{ifft}((\bar{\mathcal{S}}_k)^{\tau}_{\text{tnn}}, [], k)$ .

**Theorem 4 (Mode- $k$  LogTNN-Based t-SVT):** Assuming that  $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is a three-way tensor, a minimizer to

$$\arg \min_{\mathcal{Z}} \tau \text{LogTNN}_k(\mathcal{Z}, \varepsilon) + \frac{1}{2} \|\mathcal{Z} - \mathcal{Y}\|_F^2$$

is given by the mode- $k$  LogTNN-based t-SVT

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

where  $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\text{T}_k}$ ,  $\mathcal{S}_{\text{lt}}^{\tau, \varepsilon} = \text{ifft}((\bar{\mathcal{S}}_k)^{\tau, \varepsilon}_{\text{lt}}, [], k)$ , and

$$(\bar{\mathcal{S}}_k)^{\tau, \varepsilon}_{\text{lt}}(i, j, s) = \begin{cases} 0, & \text{if } c_2 \leq 0 \\ \text{sign}(\bar{\mathcal{S}}_k(i, j, s)) \left( \frac{c_1 + \sqrt{c_2}}{2} \right), & \text{if } c_2 > 0 \end{cases}$$

where  $\bar{\mathcal{S}}_k = \text{fft}(\mathcal{S}, [], k)$ ,  $c_1 = |(\bar{\mathcal{S}}_k(i, j, s))| - \varepsilon$ , and  $c_2 = c_1^2 - 4(\tau - \varepsilon|(\bar{\mathcal{S}}_k(i, j, s))|)$ .

Theorem 4 can be easily proved by [39, Th. 2]. Especially, according to Theorem 5, we can easily obtain that the mode- $k$  LogTNN-based t-SVT can make larger singular values shrunk less than small singular values. This implies that 3DlogTNN is able to better preserve the major information of the target HSI.

---

### Algorithm 2 ADMM Solver for the 3DTNN-Based and 3DLogTNN-Based HSI Denosing Models

---

**Input:** The noisy HSI  $\mathcal{Y}$ , parameters  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ ,  $\mu = (\mu_1, \mu_2, \mu_3)$ ,  $\lambda_1$ ,  $\lambda_2$ ,  $\beta$  and  $\rho = 1.2$ .

**Initialization:**  $p = 0$ ,  $\mathcal{X}^0 = 0$ ,  $\mathcal{N}^0 = 0$ ,  $\mathcal{S}^0 = 0$ ,  $\mathcal{Z}_k^0 = 0$ ,  $\mathcal{M}_k^0 = 0$ , and  $\mathcal{P}^0 = 0$ .

1: **while** not converged **do**

2:   Update  $\mathcal{Z}_k^{p+1}$  via (7) or (8),  $k = 1, 2, 3$ .

3:   Update  $\mathcal{X}^{p+1}$  via (10).

4:   Update  $\mathcal{N}^{p+1}$  via (12).

5:   Update  $\mathcal{S}^{p+1}$  via (14).

6:   Update  $\mathcal{M}_k^{p+1}$  and  $\mathcal{P}^{p+1}$  via (5).

7:   Let  $\mu = \rho\mu$ ;  $\beta = \rho\beta$ ;  $p = p + 1$ .

8:   Check the convergence condition:

$$\|\mathcal{X}^{(p+1)} - \mathcal{X}^{(p)}\|_F / \|\mathcal{X}^{(p)}\|_F < 10^{-4}.$$

9: **end while**

**Output:** The restored HSI  $\mathcal{X}$ .

---

**Theorem 5:** Defining the function

$$f(x) = x - \frac{c_1(x) + \sqrt{c_2(x)}}{2}$$

where  $c_1(x) = x - \varepsilon$ ,  $c_2(x) = c_1(x)^2 - 4(\tau - \varepsilon x)$ , and  $\varepsilon > 0$  and  $\tau > 0$  are constants. Then,  $f(x)$  is a monotonically decreasing function when  $x \in [0, +\infty)$ .

*Proof.* Simplifying  $f(x)$ :

$$\begin{aligned} f(x) &= x - \frac{c_1(x) + \sqrt{c_2(x)}}{2} \\ &= x - \frac{(x - \varepsilon) + \sqrt{(x - \varepsilon)^2 - 4(\tau - \varepsilon x)}}{2} \\ &= \frac{(x + \varepsilon) - \sqrt{(x + \varepsilon)^2 - 4\tau}}{2}. \end{aligned}$$

The derivative of  $f(x)$  is

$$f'(x) = \frac{1}{2} \left( 1 - \frac{x + \varepsilon}{\sqrt{(x + \varepsilon)^2 - 4\tau}} \right) < 0, \quad \text{when } x \geq 0.$$

Thus,  $f(x)$  is a monotonically decreasing function.

For the proposed 3DTNN-based HSI model, by using Theorem 3,  $\mathcal{Z}_k$  ( $k = 1, 2, 3$ ) subproblems can be solved as

$$\mathcal{Z}_k^{p+1} = \mathcal{D}_{\text{tnn}}^{\frac{\alpha_k}{\mu_k}} \left( \mathcal{X}^p + \frac{\mathcal{M}_k^p}{\mu_k}, k \right). \quad (7)$$

For the proposed 3DLogTNN-based HSI model, by using Theorem 4,  $\mathcal{Z}_k$  ( $k = 1, 2, 3$ ) subproblems can be solved as

$$\mathcal{Z}_k^{p+1} = \mathcal{D}_{\text{lt}}^{\frac{\alpha_k}{\mu_k}, \varepsilon} \left( \mathcal{X}^p + \frac{\mathcal{M}_k^p}{\mu_k}, k \right). \quad (8)$$

The complexity of computing  $\mathcal{Z}_k$  ( $k = 1, 2, 3$ ), in both 3DTNN-based solver and 3DLogTNN-based solver, is  $\mathcal{O}(n_1 n_2 n_3 (\log(n_1 n_2 n_3) + \sum_{i=1}^3 \min(n_i, n_{i+1})))$ , where we define  $n_4 = n_1$ .

The  $\mathcal{X}$  subproblem is

$$\mathcal{X}^{p+1} = \arg \min_{\mathcal{X}} \sum_{k=1}^3 \frac{\mu_k}{2} \left\| \mathcal{X} - \mathcal{Z}_k^{p+1} + \frac{\mathcal{M}_k^p}{\mu_k} \right\|_F^2 + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X} + \mathcal{N}^p + \mathcal{S}^p) + \frac{\mathcal{P}^p}{\beta} \right\|_F^2 \quad (9)$$

which has the following closed-form solution:

$$\mathcal{X}^{p+1} = \frac{\sum_{k=1}^3 \mu_k \left( \mathcal{Z}_k^{p+1} - \frac{\mathcal{M}_k^p}{\mu_k} \right) + \beta \left( \mathcal{Y} - \mathcal{N}^p - \mathcal{S}^p + \frac{\mathcal{P}^p}{\beta} \right)}{\sum_{k=1}^3 \mu_k + \beta}. \quad (10)$$

The complexity of computing  $\mathcal{X}$  is  $\mathcal{O}(n_1 n_2 n_3)$ .

The  $\mathcal{N}$  subproblem is

$$\mathcal{N}^{p+1} = \arg \min_{\mathcal{N}} \lambda_1 \|\mathcal{N}\|_F^2 + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X}^{p+1} + \mathcal{N} + \mathcal{S}^p) + \frac{\mathcal{P}^p}{\beta} \right\|_F^2 \quad (11)$$

which has the following closed-form solution:

$$\mathcal{N}^{p+1} = \frac{\beta \left( \mathcal{Y} - \mathcal{X}^{p+1} - \mathcal{S}^p + \frac{\mathcal{P}^p}{\beta} \right)}{2\lambda_1 + \beta}. \quad (12)$$

The complexity of computing  $\mathcal{N}$  is  $\mathcal{O}(n_1 n_2 n_3)$ .

The  $\mathcal{S}$  subproblem is

$$\mathcal{S}^{p+1} = \arg \min_{\mathcal{S}} \lambda_2 \|\mathcal{S}\|_1 + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X}^{p+1} + \mathcal{N}^{p+1} + \mathcal{S}) + \frac{\mathcal{P}^p}{\beta} \right\|_F^2 \quad (13)$$

which has the following solution:

$$\mathcal{S}^{p+1} = \text{shrink} \left( \mathcal{Y} - \mathcal{X}^{p+1} - \mathcal{N}^{p+1} + \frac{\mathcal{P}^p}{\beta}, \frac{\lambda_2}{\beta} \right) \quad (14)$$

where  $\text{shrink}(\cdot, \zeta)$  is the tensor soft thresholding operator with threshold  $\zeta$ , that is,

$$[\text{shrink}(\mathcal{X}, \zeta)]_{ijs} = \text{sgn}(x_{ijs}) \max(|x_{ijs}| - \zeta, 0). \quad (15)$$

The complexity of computing  $\mathcal{S}$  is  $\mathcal{O}(n_1 n_2 n_3)$ .

The computational cost at each iteration of the proposed algorithms are  $\mathcal{O}(n_1 n_2 n_3 (\log(n_1 n_2 n_3) + \sum_{i=1}^3 \min(n_i, n_{i+1})))$ , where we define  $n_4 = n_1$ . The pseudocode of the developed ADMM-based solving algorithm is described in Algorithm 2.

## V. NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of the proposed 3DTNN and 3DLogTNN-based HSI denoising models on both simulated and real-world HSIs. All the testing HSIs are normalized to [0,1] band by band. The comparison methods are listed below.

- 1) TRPCA+BM4D [36], [58] uses TRPCA to filter sparse noise and then performs BM4D to remove Gaussian noise.
- 2) LRMR [43] separates the target HSI as a series of full-band cubes, unfolds each full-band cube as a matrix by

TABLE II

PARAMETERS SETTING IN THE PROPOSED 3DTNN-BASED AND 3DLogTNN-Based HSI Denoising Methods for Simulated Data

Method	Data	Case	$\alpha$	$\tau$	$\lambda_1$	$\lambda_2$	$\varepsilon$
3DTNN	Washington DC Mall	Case 1	$\frac{(1,1,0.001)}{2.001}$	$(10^2, 10^2, 10^2)$	$\frac{0.002}{\sigma}$	$\lambda$	—
		Case 2				$\lambda$	
		Case 3				$\lambda$	
		Case 4				$0.7\lambda$	
	Pavia City Center	Case 1	$\frac{(1,1,0.001)}{2.001}$	$(10^4, 10^4, 10^4)$	$\frac{0.004}{\sigma}$	$\lambda$	60
		Case 2				$\lambda$	
		Case 3				$\lambda$	
3DLogTNN	Washington DC Mall	Case 1	$\frac{(1,1,0.001)}{2.001}$	$(10^4, 10^4, 10^4)$	$\frac{0.00005}{\sigma}$	$0.011\lambda$	60
		Case 2				$0.011\lambda$	
		Case 3				$0.011\lambda$	
		Case 4				$0.007\lambda$	
	Pavia City Center	Case 1	$\frac{(1,1,0.001)}{2.001}$	$(10^4, 10^4, 10^4)$	$\frac{0.00005}{\sigma}$	$0.011\lambda$	80
		Case 2				$0.011\lambda$	
		Case 3				$0.011\lambda$	

vectorizing each band as a column, and then promotes the low rankness of the unfolding matrix.

- 3) LRTR [60] formulates the target HSI as a low-tubal-rank component and minimizes the tensor nuclear norm, a convex relaxation of the tubal rank, of the target HSI.
- 4) LRTDTV [29] uses Tucker decomposition to exploit the global correlation in an HSI and employs SSTV to describe the piecewise smooth prior along both spatial and spectral modes.
- 5) NMoG [55] employs LRMF to promote the low rankness of the target HSI and uses a non-i.i.d. mixture of Gaussian noise assumption to model the noise.

Parameters of all methods are set based on authors' codes or suggestions in their articles. All tests are implemented on the platform of Windows 10 and MATLAB (R2018b) with an Intel Core i9-9900K 3.60 GHz and 32 GB RAM.

### A. Mixed Noise Removal on Simulated Data

In this section, we adopt two HSIs, i.e., a subimage of *Washington DC Mall* data set<sup>2</sup> of size  $256 \times 256 \times 191$  and a subimage of *Pavia City Center* data set<sup>3</sup> of size  $200 \times 200 \times 80$  as the testing data. We employ the mean of peak signal-to-noise rate (MPSNR) over all bands, the mean of structural similarity (MSSIM) over all bands, and the spectral angle mapping (SAM) to measure the quality of the recovered results.

Real HSIs are usually degraded by a mix of various noises. Thus, to simulate real-noise scenarios, we consider different combinations of Gaussian noise with different standard deviations  $\sigma$ , salt and pepper noise with different intensities  $v$ , and stripes with different proportions  $s$ .

*Case 1 i.i.d. Gaussian Noise With Different Intensities + i.i.d. Salt and Pepper Noise:* This case includes three subcases. All bands are corrupted by i.i.d. Gaussian noise with zero mean and standard deviation  $\sigma = 0.05, 0.10, 0.15$ . In addition,

<sup>2</sup><http://lesun.weebly.com/hyperspectral-data-set.html>.

<sup>3</sup>[http://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).



TABLE III  
PERFORMANCE COMPARISON OF SEVEN COMPETING METHODS WITH RESPECT TO DIFFERENT NOISE LEVELS

Case	Data	Gaussian noise	Salt and pepper noise	Stripes	Indicators	Noise	TRPCA +BM4D	LRMR	LRTR	LRTDTV	NMoG	3DTNN	3DLogTNN
Case 1	Washington DC Mall	$\sigma = 0.05$	$v = 0.20$	—	MPSNR	11.232	34.716	34.297	32.591	36.161	39.497	37.884	39.515
					MSSIM	0.1158	0.9486	0.9376	0.9206	0.9567	0.9770	0.9767	0.9815
					SAM	47.797	5.2919	6.3661	7.5319	3.8320	3.1288	3.2314	2.7423
		$\sigma = 0.10$			MPSNR	10.826	31.032	31.608	29.141	32.968	34.728	33.242	35.654
					MSSIM	0.1023	0.8837	0.8960	0.8319	0.9155	0.9414	0.9376	0.9562
					SAM	49.169	7.3313	8.6401	10.182	5.6550	5.5936	5.3416	4.1582
		$\sigma = 0.15$			MPSNR	10.212	28.814	29.261	26.895	29.724	31.974	29.965	32.971
					MSSIM	0.0865	0.8220	0.8436	0.7917	0.8567	0.8976	0.8776	0.9276
					SAM	51.166	8.8980	10.871	12.048	7.7469	7.6360	5.3461	
		Average time (s)					—	722.65	67.295	70.568	226.51	186.65	352.17
	Pavia City Center	$\sigma = 0.05$	$v = 0.20$	—	MPSNR	11.650	35.044	32.251	32.276	35.994	38.006	35.833	37.270
					MSSIM	0.1193	0.9555	0.9122	0.9172	0.9497	0.9733	0.9693	0.9743
					SAM	41.681	3.5273	5.5659	5.6869	4.5189	2.7973	2.4665	2.3184
		$\sigma = 0.10$			MPSNR	11.197	31.336	29.934	28.688	31.946	32.495	31.938	33.644
MSSIM					0.1048	0.8988	0.8620	0.8346	0.8961	0.9263	0.9277	0.9434	
SAM					43.261	4.4055	7.0425	7.3273	5.8839	4.4004	3.4554	3.1641	
$\sigma = 0.15$		MPSNR			10.530	29.003	27.862	26.612	28.503	29.440	29.592	31.498	
	MSSIM	0.0873	0.8355	0.7987	0.7601	0.8291	0.8733	0.8811	0.9118				
Average time (s)					—	155.43	23.687	15.029	66.113	66.241	73.450	102.14	
Case 2	Washington DC Mall	$\sigma = 0.10$	$v = 0.10$	—	MPSNR	13.330	31.853	33.180	30.479	33.968	35.957	34.701	36.578
					MSSIM	0.1816	0.8998	0.9193	0.8897	0.9301	0.9496	0.9551	0.9634
					SAM	45.035	6.6428	7.4550	9.0853	4.8662	4.9520	4.3752	3.7712
			$v = 0.30$		MPSNR	9.2348	29.930	29.909	27.041	31.812	33.435	29.913	34.507
					MSSIM	0.0648	0.8607	0.8652	0.7702	0.8940	0.9301	0.8509	0.9460
					SAM	50.922	8.3004	9.9656	12.411	6.6686	6.4081	8.7064	4.7035
		Average time (s)					—	722.14	66.356	77.449	225.68	166.56	369.43
	Pavia City Center	$\sigma = 0.10$	$v = 0.10$	—	MPSNR	13.672	32.168	31.320	30.176	33.316	33.714	32.848	34.403
					MSSIM	0.1871	0.9130	0.8929	0.8830	0.9231	0.9374	0.9409	0.9517
					SAM	38.435	4.0893	6.3813	6.2367	4.9273	4.0363	3.1310	2.9704
			$v = 0.30$		MPSNR	9.6252	30.206	28.503	26.503	30.472	31.057	30.821	32.912
					MSSIM	0.0657	0.8764	0.8221	0.7346	0.8598	0.9100	0.9050	0.9330
					SAM	45.570	4.8623	7.7006	9.5495	7.1364	4.8549	4.1122	3.3251
		Average time (s)					—	155.01	24.211	15.896	66.188	65.009	73.790
Case 3	Washington DC Mall	$\sigma$ follows $U(0.05, 0.15)$	$v = 0.20$	—	MPSNR	10.803	31.197	31.562	29.137	33.029	35.655	33.253	35.921
					MSSIM	0.1024	0.8873	0.8948	0.8370	0.9157	0.9460	0.9368	0.9599
					SAM	49.241	7.2224	8.7228	10.570	5.6648	7.3789	5.5289	4.1032
		$\sigma$ follows $U(0.10, 0.20)$			MPSNR	10.136	28.748	29.087	26.552	29.580	32.202	29.494	32.813
					MSSIM	0.0853	0.8196	0.8394	0.7424	0.8530	0.9015	0.8557	0.9281
					SAM	51.461	8.9559	11.063	13.223	7.7146	7.5588	8.4352	5.5482
		Average time (s)					—	731.55	69.651	74.142	225.90	229.24	374.30
	Pavia City Center	$\sigma$ follows $U(0.05, 0.15)$	$v = 0.20$	—	MPSNR	11.127	31.246	29.640	28.398	31.711	32.864	31.879	33.734
					MSSIM	0.1024	0.8967	0.8527	0.8210	0.8861	0.9194	0.9267	0.9453
					SAM	43.568	4.3957	7.4626	7.6978	6.8542	8.4942	3.6976	3.1680
		$\sigma$ follows $U(0.10, 0.20)$			MPSNR	10.496	29.031	27.734	26.491	28.506	28.876	29.543	31.501
					MSSIM	0.0866	0.8355	0.7944	0.7481	0.8238	0.8394	0.8779	0.9120
					SAM	45.972	5.0151	8.6105	8.6141	7.1666	13.152	4.5618	3.8384
		Average time (s)					—	157.04	24.401	14.455	65.419	73.183	76.957
Case 4	Washington DC Mall	$\sigma$ follows $U(0.10, 0.20)$	$v$ follows $U(0.10, 0.30)$	$s = 10\%$	MPSNR	10.496	29.014	29.400	27.090	29.855	31.978	29.255	31.636
					MSSIM	0.0943	0.8234	0.8476	0.7681	0.8572	0.8914	0.8718	0.9109
					SAM	51.015	8.6708	10.828	12.480	7.5715	15.858	7.8254	6.2912
				$s = 20\%$	MPSNR	10.302	28.683	28.959	26.618	29.474	31.196	28.805	31.067
					MSSIM	0.0897	0.8147	0.8371	0.7428	0.8477	0.8741	0.8604	0.9026
					SAM	51.397	9.0045	11.228	13.094	8.0032	17.733	8.2952	6.8157
		Average time (s)					—	746.98	72.351	64.323	231.44	211.96	324.13

all bands are corrupted by i.i.d. salt and pepper noise with proportion  $v = 0.20$ .

*Case 2 i.i.d. Gaussian Noise + i.i.d. Salt and Pepper Noise With Different Intensities:* This case includes two subcases. All bands are corrupted by i.i.d. Gaussian noise with zero mean and standard deviation  $\sigma = 0.10$ . In addition, all bands are corrupted by i.i.d. salt and pepper noise with proportions  $v = 0.10, 0.30$ .

*Case 3 Non-i.i.d. Gaussian Noise With Different Intensities + i.i.d. Salt and Pepper Noise:* This case includes two

subcases. All bands are corrupted by non-i.i.d. Gaussian noise with zero mean and bandwise standard deviation  $\sigma$  randomly sampled from uniform distribution  $U(0.05, 0.15)$  and  $U(0.10, 0.20)$ . In addition, all bands are corrupted by i.i.d. salt and pepper noise with proportion  $v = 0.20$ .

*Case 4 Non-i.i.d. Gaussian Noise + Non-i.i.d. Salt and Pepper Noise + Stripes With Different Proportions:* This case includes two subcases. All bands are corrupted by non-i.i.d. Gaussian noise with zero mean and bandwise standard deviation  $\sigma$  randomly sampled from uniform distribution

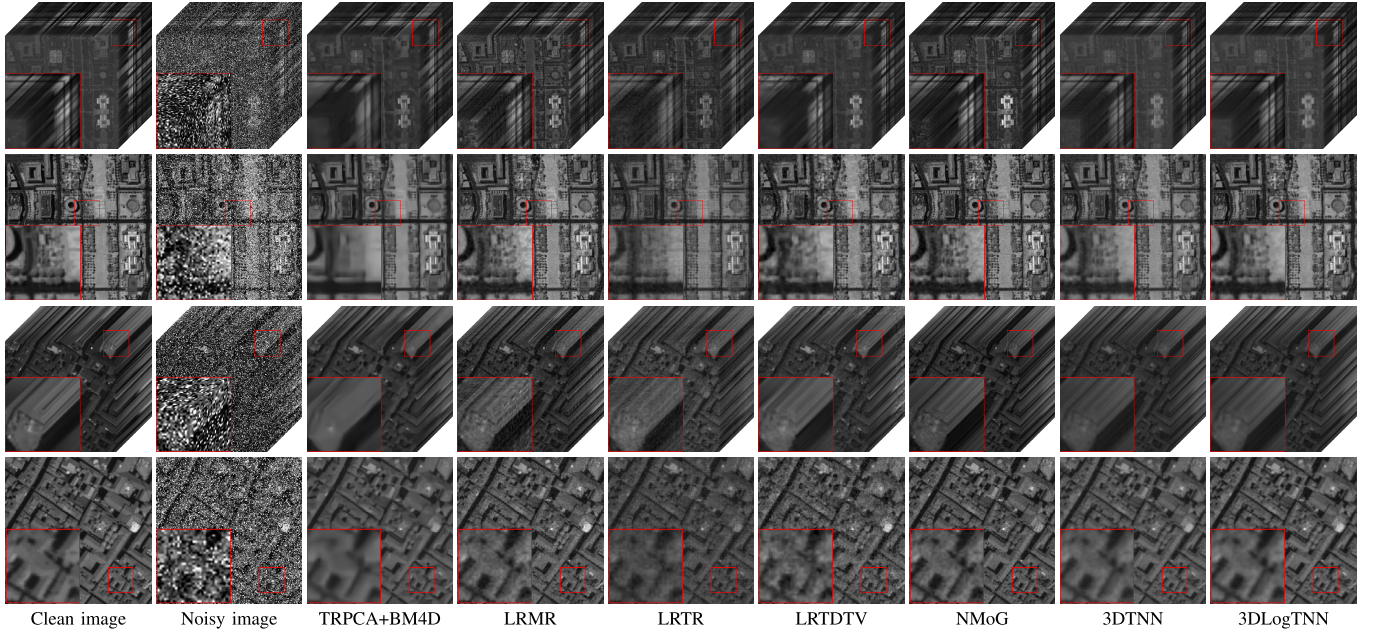


Fig. 5. Case 1: standard deviation  $\sigma$  of the Gaussian noise is 0.15 and the proportion  $\nu$  of the salt and pepper noise is 0.20. (From Top to Bottom) 3-D visualization of the denoising results for the HSI *Washington DC Mall*, the denoising results at band 71 of the HSI *Washington DC Mall*, the 3-D visualization of the denoising results for the HSI *Pavia City Center*, and the denoising results at band 80 of the HSI *Pavia City Center*.

$U(0.10, 0.20)$ . All bands are corrupted by non-i.i.d. salt and pepper noise and bandwise proportion  $\nu$  randomly sampled from uniform distribution  $U(0.10, 0.30)$ . Ten bands from band 131 to 140 of the HSI *Washington DC Mall* are corrupted by stripes with proportion  $s = 10\%$ ,  $20\%$ .

The proposed models involve the weight  $\alpha$ , the regularization parameters  $\lambda_1$  and  $\lambda_2$ , and a constant  $\varepsilon$  for 3DLogTNN; the ADMM solver (5) involves the threshold parameter  $\tau = \alpha/\mu$  and the penalty parameter  $\beta$ . We adjust these parameters to achieve the highest PSNR value. In all the tests, we set the parameter  $\beta$  to  $1/\text{mean}(\tau)$  and the other parameters setting are reported in Table II, where

$$\lambda = \frac{\alpha_1}{\sqrt{\max(n_2, n_3)n_1}} + \frac{\alpha_2}{\sqrt{\max(n_3, n_1)n_2}} + \frac{\alpha_3}{\sqrt{\max(n_1, n_2)n_3}}.$$

All parameters setting are referring to their initial value. More detailed parameter analysis will be discussed in Section V-C.

Table III lists the MPSNR, MSSIM, SAM, and average running time (in seconds) of seven competing methods on HSIs *Washington DC Mall* and *Pavia City Center*. The best and the second-best results are highlighted by bold and underline, respectively. We observe that the 3DLogTNN-based method has an overall better performance than the compared methods for both data sets and all noise cases. Although in rare cases, the 3DLogTNN-based method obtains slightly lower MPSNR than NMoG, it outperforms NMoG in terms of MSSIM and SAM. The proposed 3DTNN-based method has a better overall performance than the TRPCA+BM4D, LRMR, LRTR, and LRTDTV. And in most cases, the 3DTNN-based method underperforms NMoG in term of MPSNR, while outperforming NMoG in term of SAM. Although the proposed methods are not the fastest ones, their running times are acceptable, considering that the proposed methods achieve much better results.

We further visually compare the performance of seven competing methods. For cases 1–3, we select two noise settings and display the denoising results of the testing HSIs in Figs. 5 and 6, respectively. As observed, the proposed methods produce visually superior results than the compared methods. Specifically, the proposed methods are capable of better removing the Gaussian noise and salt and pepper noise while finely preserving the structure of the underlying HSIs, while the results obtained by LRMR, LRTR, and NMoG remain a small amount of noise. The TRPCA+BM4D and LRTDTV can perform comparatively better in noise removing, but their results contain some evident blurry area, leading to some details missing. Especially, the proposed 3DLogTNN-based method is superior to the 3DTNN-based method, in the recovery of both shape structure and texture information. The main reason is that 3DLogTNN can better approximate the tensor fibered rank than 3DTNN.

For case 4, it mainly tests the performance for stripes removal. Fig. 7 shows the denoising results at band 131. As observed, the proposed 3DLogTNN-based HSI denoising method is capable of better removing the unexpected stripes while finely preserving the details of the underlying HSI, while the results obtained by LRTR and the proposed 3DTNN-based method ( $s = 20\%$ ) contain a small number of stripes. The TRPCA+BM4D and LRTDTV are able to remove all noises better, but leading to some details missing. The LRMR and NMoG can perform better in stripes removal, but their results remain a small amount of other noise.

#### B. Mixed Noise Removal on Real Data

In this section, we adopt two HSIs, i.e., a subimage of *Indian Pines* data set<sup>4</sup> of size  $128 \times 128 \times 220$  and a subimage

<sup>4</sup><https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>.



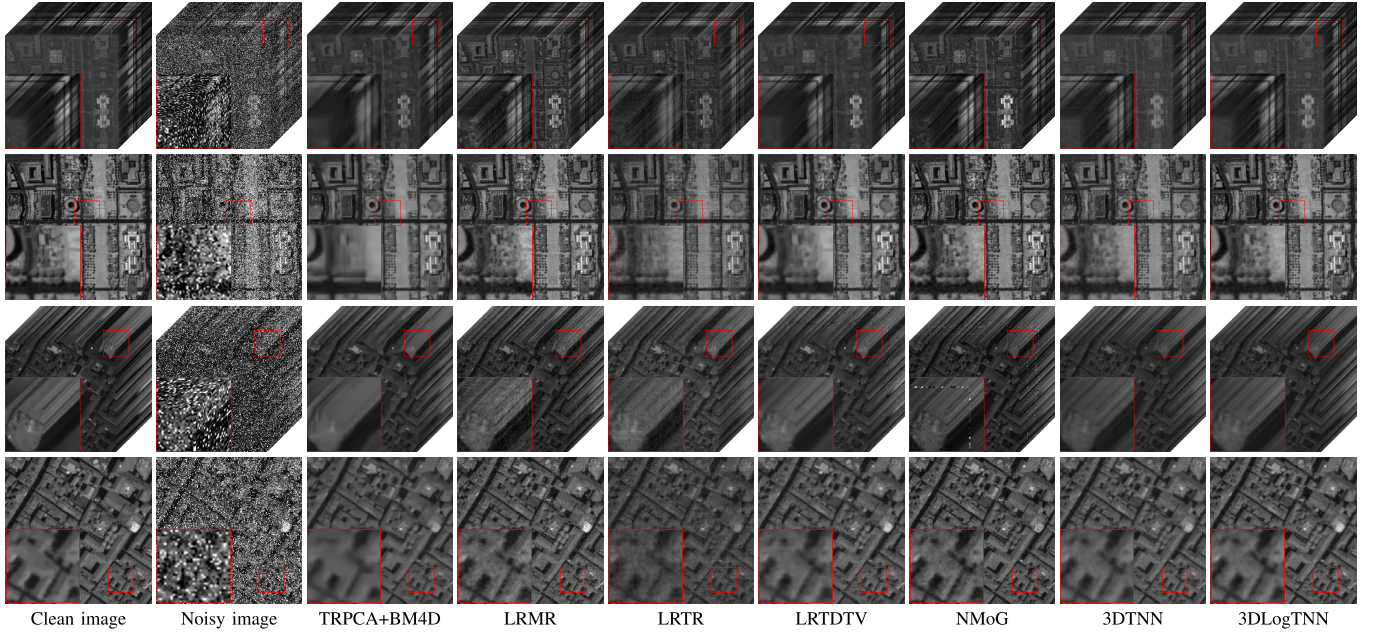


Fig. 6. Case 3: standard deviation  $\sigma$  of the Gaussian noise in each band follows  $U(0.10, 0.20)$  and the proportion  $v$  of the salt and pepper noise is 0.20. (From Top to Bottom) 3-D visualization of the denoising results for the HSI *Washington DC Mall*, the denoising results at band 71 of the HSI *Washington DC Mall*, the 3-D visualization of the denoising results for the HSI *Pavia City Center*, and the denoising results at band 80 of the HSI *Pavia City Center*.

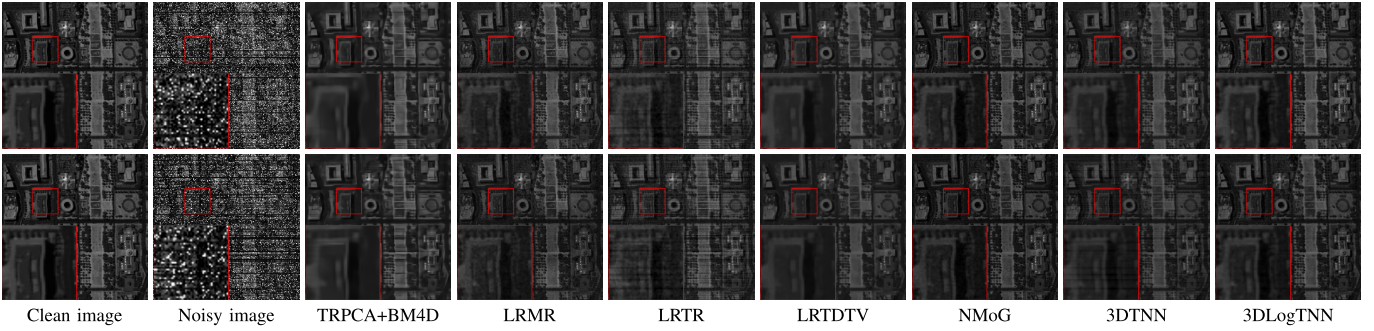


Fig. 7. Case 4: denoising results at band 131 of the HSI *Washington DC Mall*. The standard deviation  $\sigma$  of the Gaussian noise in each band follows  $U(0.05, 0.15)$ , the proportion  $v$  of the salt and pepper noise in each band follows  $U(0.10, 0.30)$ , and the percent  $s$  of the stripe noise are 10% and 20%, respectively.

of *Australian* data set<sup>5</sup> of size  $200 \times 200 \times 150$  as the testing data. Especially, the HSI *Indian Pines* mainly contains the Gaussian noise and the sparse noise and the HSI *Australian* mainly contains the stripe noise. We adjust the parameters to achieve the best visual result, and the parameter setting are reported in Table IV.

Fig. 8 shows the visual results and the corresponding vertical mean profiles of band 103 in HSI *Indian Pines* and band 50 in HSI *Australian*. We observe that the observed images are affected by heavy noise, leading to rapid fluctuations in the vertical mean profiles. The proposed 3DTNN-based and 3DLogTNN-based methods are able to produce superior results than the compared ones. Especially, the proposed methods are capable of better removing the mixed noise, only retaining a small number of inconspicuous stripes on *Australian*, and the corresponding mean profiles reduce fluctuations caused by noises and follow the basic trend of

TABLE IV  
PARAMETERS SETTING IN THE PROPOSED 3DTNN-BASED AND 3DLOGTNN-BASED HSI DENOISING METHODS FOR REAL DATA

Method	Data	$\alpha$	$\tau$	$\lambda_1$	$\lambda_2$	$\varepsilon$
3DTNN	<i>Indian Pines</i>	$(\frac{1,1,0.001}{2.001})$	$(10^2, 10^2, 10^2)$	0.05	$0.5\lambda$	—
	<i>Australian</i>	$(\frac{1,1,0.001}{2.001})$	$(10^2, 10^2, 10^2)$	0.05	$0.5\lambda$	—
3DLogTNN	<i>Indian Pines</i>	$(\frac{1,1,0.001}{2.001})$	$(10^4, 10^4, 10^4)$	0.001	$0.005\lambda$	50
	<i>Australian</i>	$(\frac{1,1,0.001}{2.001})$	$(10^4, 10^4, 10^4)$	0.001	$0.005\lambda$	50

the observed profiles. In comparison, the BM4D+TRPCA, LRTR, and LRTDTV produce mean profiles deviating from the basic trend of the observed profile on *Indian Pines*, and their results on *Australian* contain obvious noises, resulting in rapid fluctuations in corresponding mean profiles. LRMR and NMoG perform comparatively better on HSI *Indian Pines*, but their results on HSI *Australian* contain a large number of noises, and the mean profiles of them are still rapid fluctuations such as the observed one.

<sup>5</sup><http://remote-sensing.nci.org.au/u39/public/html/index.shtml>.

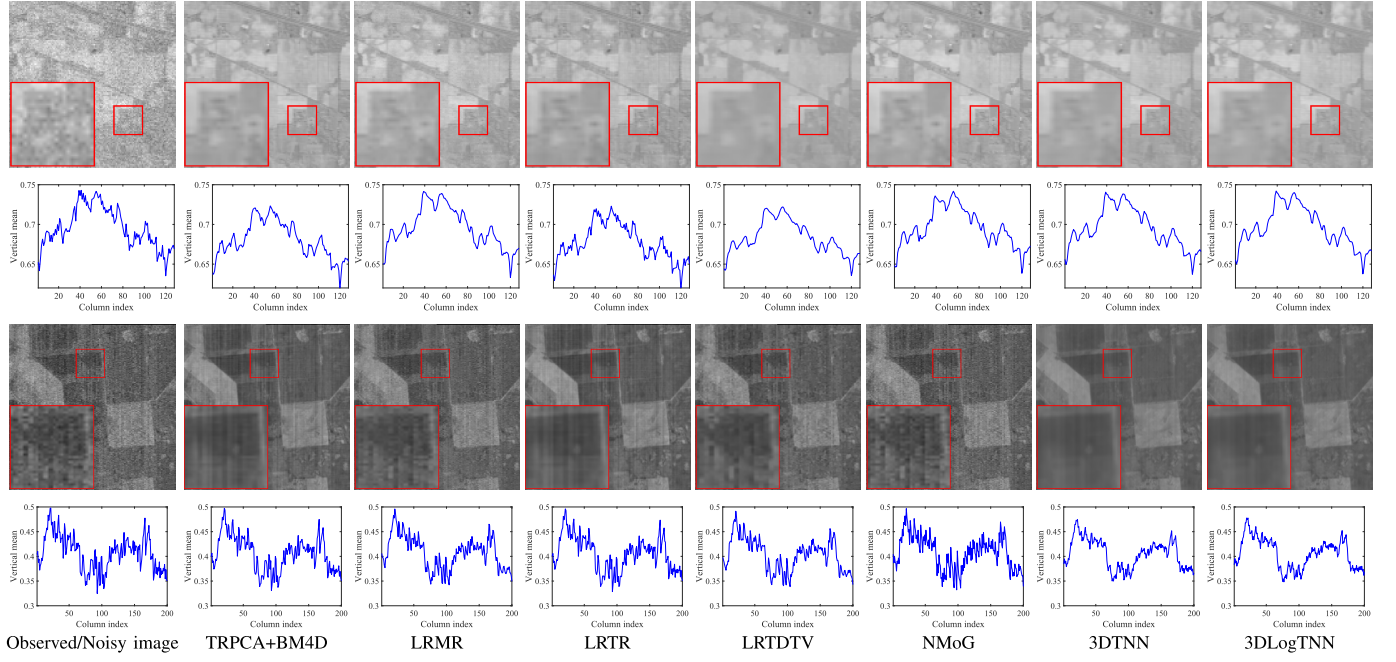


Fig. 8. Denoising results of the real HSIs *Indian Pines* and *Australian*. The top two rows are the visual results and the vertical mean profiles of band 103 in *Indian Pines*, respectively. The bottom two rows are the visual results and the vertical mean profiles of band 50 in *Australian*, respectively.

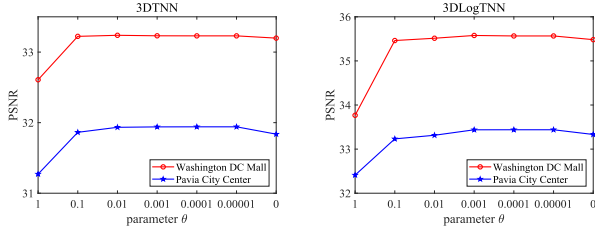


Fig. 9. PSNR values with respect to different values of  $\theta$  (controls  $\alpha$ ).

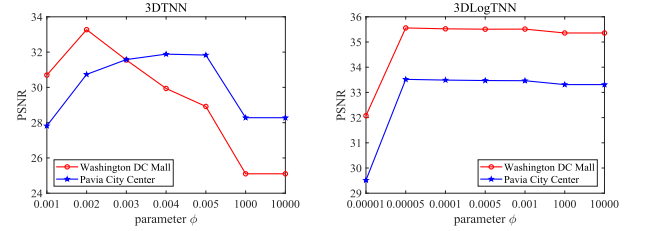


Fig. 10. PSNR values with respect to different values of  $\phi$  (controls  $\lambda_1$ ).

### C. Discussions

1) *Parameter Analysis*: We analyze the robustness of the parameters involved in the proposed method on HSIs *Washington DC Mall* and *Pavia City Center*, i.e., the weight  $\alpha$ , the regularization parameters  $\lambda_1$  and  $\lambda_2$ , the threshold parameter  $\tau = \alpha/\mu$ , the penalty parameter  $\beta$ , and a constant  $\varepsilon$  in 3DLogTNN. The standard deviation  $\sigma$  of Gaussian noise is 0.10 and the proportion  $\nu$  of salt and pepper noise is 0.20.

The weight  $\alpha$  controls the weight of the correlation along each mode of HSIs. Since the correlation along the spectral mode of HSIs should be much stronger than those along spatial mode, we empirically choose the weights  $\alpha$  as  $(1, 1, \theta)/(2+\theta)$ , where  $0 \leq \theta \leq 1$  is a balance parameter. The robustness analysis of  $\theta$  is shown in Fig. 9. As observed, when  $\theta \leq 0.1$ , the proposed methods achieve nearly identical PSNR values. Especially, when  $\theta = 0$ , i.e.,  $\alpha = (1/2, 1/2, 0)$ , the PSNR value is about 0.1 dB lower than the best results.

The regularization parameter  $\lambda_1$  controls the weight of Gaussian noise term, which is set to  $\phi/\sigma$ . The robustness analysis of  $\phi$  is shown in Fig. 10. For the 3DTNN-based method, we observe that its performance is sensitive to  $\phi$ . Especially, the highest PSNR is achieved at  $\phi = 0.002$  for the

HSI *Washington DC Mall* and  $\phi = 0.004$  for the HSI *Pavia City Center*. For the 3DLogTNN-based method, the PSNR curves are stable for  $\phi \geq 0.00005$ . Especially, we test the case of very large  $\phi$  ( $\phi = 10000$ ), in which case the Gaussian noise term  $\|\mathcal{N}\|_F^2$  tends to 0, i.e., the proposed model degenerates to the TRPCA model

$$\min_{\mathcal{X}, \mathcal{S}} \text{rank}_f(\mathcal{X}) + \lambda_2 \|\mathcal{S}\|_1, \quad \text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{S}.$$

It can be observed from Fig. 10 that in this case, the performance of the 3DTNN-based method is greatly degraded, while that of the 3DLogTNN-based method is only slightly affected.

The regularization parameter  $\lambda_2$  controls the weight of sparse noise term. Defining

$$\lambda = \frac{\alpha_1}{\sqrt{\max(n_2, n_3)n_1}} + \frac{\alpha_2}{\sqrt{\max(n_3, n_1)n_2}} + \frac{\alpha_3}{\sqrt{\max(n_1, n_2)n_3}}$$

we set  $\lambda_2$  to  $\varpi\lambda$ . The robustness analysis of  $\varpi$  is shown in Fig. 11. As observed, both the 3DTNN and the 3DLogTNN-based methods are sensitive to  $\varpi$ . Specially, the 3DTNN-based method obtains the highest PSNR value when  $\varpi = 1$  and the 3DLogTNN-based method achieves the highest PSNR value when  $\varpi = 0.011$ .



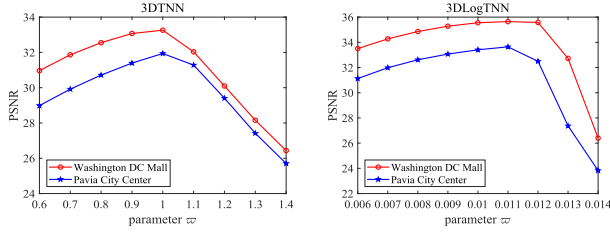


Fig. 11. PSNR values with respect to different values of  $\omega$  (controls  $\lambda_2$ ).

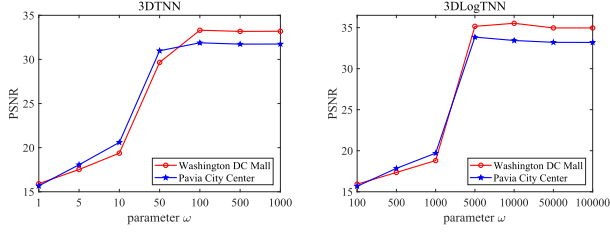


Fig. 12. PSNR values with respect to different values of  $\omega$  (controls  $\tau$ ).

The threshold parameter  $\tau$  and the penalty parameter  $\beta$  are used in the ADMM solver. The parameter  $\tau$  is set to  $\omega \times (1, 1, 1)$  and the parameter  $\beta$  is set to  $1/\text{mean}(\tau)$ . The robustness analysis of  $\omega$  is shown in Fig. 12. We observe that small values of  $\tau$  lead to poor results. The main reason is that too little shrinkage of singular values results in insufficient noise removal. Considering that too large shrinkage of singular values result in detail missing,  $\omega$  should be selected from a moderate range.

The constant  $\varepsilon$  is involved in 3DLogTNN. The robustness analysis of  $\varepsilon$  is presented in Fig. 13. As observed, too large or too small  $\varepsilon$  output unfavorable results, while the moderate values obtain the best results. Under the guidance of this observation, the parameter  $\varepsilon$  is set to 60 in the tests on the HSI *Washington DC Mall* and set to 80 in the tests on the HSI *Pavia City Center*.

2) *Convergence Analysis*: We analyze the convergence of the proposed ADMM-based algorithm on the testing HSIs *Washington DC Mall* and *Pavia City Center*. The standard deviation  $\sigma$  of Gaussian noise is set to 0.10. The proportion of salt and pepper noise is set to 0.20.

For the proposed 3DTNN-based method, since the objective function is a convex function, the convergence of the developed algorithm within the ADMM framework is guaranteed theoretically. To illustrate the convergence numerically, in Fig. 14, we display the RelCha, i.e., the relative change of two successive reconstructed tensors, with respect to the iteration. From Fig. 14(a), we observe that as the number of iterations increases, the RelCha values obtained by the 3DTNN-based method gradually tend to zero, which clearly justifies the numerical convergence of the proposed 3DTNN-based method.

For the 3DLogTNN-based method, unfortunately, the theoretical convergence of the ADMM for a general nonconvex problem is still an open problem. Instead, we give an empirical analysis of the convergence. It can be observed from Fig. 14(b) that the RelCha values obtained by the 3DLogTNN-based

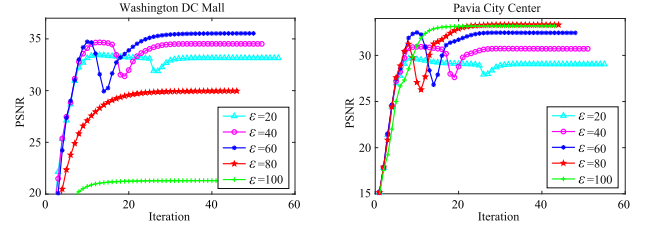


Fig. 13. PSNR values with respect to the iteration for different values of  $\varepsilon$ .

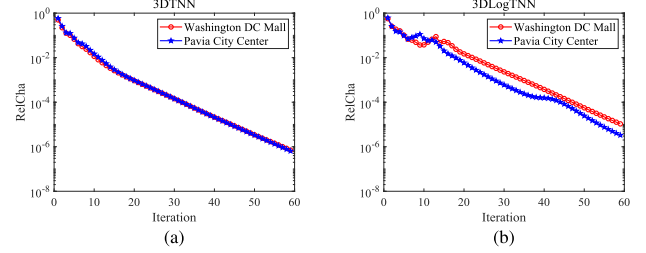


Fig. 14. RelCha values with respect to the iteration. (a) Result of the 3DTNN-based method. (b) Result of the 3DLogTNN-based method.

method are monotonically decreasing and rapidly tending to zeros after 15 iterations, although there is a slight fluctuation in the first 15 iterations. It implies the strong convergence of the 3DLogTNN-based method within the ADMM framework. Furthermore, we observe from Fig. 14 that the 3DLogTNN-based method converges slightly slower than the 3DTNN-based method on both *Washington DC Mall* and *Pavia City Center*. But considering that the 3DLogTNN-based method achieves much better denoising results, its convergence rate is acceptable.

## VI. CONCLUSION

In this article, we proposed a novel HSI mixed noise removal model by generalizing the t-SVD to the mode- $k$  t-SVD and modeling the HSI as the corresponding low-fibered-rank component. Noted that the fibered rank minimization problem is NP-hard, we introduced a 3DTNN and a 3DLogTNN as its convex and nonconvex relaxations, respectively. In particular, we showed that the 3DLogTNN is a closer approximation to the proposed fibered rank than the 3DTNN and can preserve major information by treating singular values differently. To solve the proposed model, we developed an efficient ADMM-based algorithm with empirical convergence analysis. By comparing with the state-of-the-art competing methods on low-rank matrix/tensor approximation and noise modeling, including TRPCA+BM4D [36], [58], LRMR [43], LRTR [60], LRTDTV [29], and NMoG [55], extensive experimental results demonstrated that the proposed method has superiorities of better removing the mixed noise and finely preserving the inherent structure.

## REFERENCES

- [1] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

- [2] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.
- [3] G. Wang, Y. Zhang, B. He, and K. T. Chong, "A framework of target detection in hyperspectral imagery based on blind source extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 835–844, Feb. 2016.
- [4] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.
- [5] Y. Liu, G. Gao, and Y. Gu, "Tensor matched subspace detector for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1967–1974, Apr. 2017.
- [6] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.
- [7] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
- [8] S. Prasad, D. Labate, M. Cui, and Y. Zhang, "Morphologically decoupled structured sparsity for rotation-invariant hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4355–4366, Aug. 2017.
- [9] S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1176–1188, Apr. 2018.
- [10] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013.
- [11] X. Fu, W.-K. Ma, T.-H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1128–1141, Sep. 2015.
- [12] X. Fu, W.-K. Ma, J. M. Bioucas-Dias, and T.-H. Chan, "Semiblind hyperspectral unmixing in the presence of spectral library mismatches," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5171–5184, Sep. 2016.
- [13] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.
- [14] X.-L. Zhao, F. Wang, and M. K. Ng, "A new convex optimization model for multiplicative noise and blur removal," *SIAM J. Imag. Sci.*, vol. 7, no. 1, pp. 456–475, 2014.
- [15] Y. Chang, L. Yan, H. Fang, and C. Luo, "Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1852–1866, Jun. 2015.
- [16] Y. Chang, L. Yan, T. Wu, and S. Zhong, "Remote sensing image stripe noise removal: From image decomposition perspective," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7018–7031, Dec. 2016.
- [17] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, T.-Y. Ji, and L.-J. Deng, "Matrix factorization for low-rank tensor completion using framelet prior," *Inf. Sci.*, vol. 436, pp. 403–417, Apr. 2018.
- [18] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
- [19] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, and Y. Wan, "Fast-derain: A novel video rain streak removal method using directional gradient priors," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2089–2102, Apr. 2019.
- [20] 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.
- [21] M. Ding, T.-Z. Huang, S. Wang, J.-J. Mei, and X.-L. Zhao, "Total variation with overlapping group sparsity for deblurring images under cauchy noise," *Appl. Math. Comput.*, vol. 341, pp. 128–147, Jan. 2019.
- [22] J.-H. Yang, X.-L. Zhao, J.-J. Mei, S. Wang, T.-H. Ma, and T.-Z. Huang, "Total variation and high-order total variation adaptive model for restoring blurred images with cauchy noise," *Comput. Math. Appl.*, vol. 77, no. 5, pp. 1255–1272, Mar. 2019.
- [23] J.-J. Mei, Y. Dong, T.-Z. Huang, and W. Yin, "Cauchy noise removal by nonconvex ADMM with convergence guarantees," *J. Sci. Comput.*, vol. 74, no. 2, pp. 743–766, Feb. 2018.
- [24] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 6868–6877.
- [25] X.-T. Li, X.-L. Zhao, T.-X. Jiang, Y.-B. Zheng, T.-Y. Ji, and T.-Z. Huang, "Low-rank tensor completion via combined non-local self-similarity and low-rank regularization," *Neurocomputing*, to be published. doi: [10.1016/j.neucom.2019.07.092](https://doi.org/10.1016/j.neucom.2019.07.092).
- [26] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *J. Comput. Appl. Math.*, vol. 363, pp. 124–144, Jan. 2020.
- [27] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.
- [28] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.
- [29] Y. Wang, J. Peng, Q. Zhao, D. Meng, Y. Leung, and X.-L. Zhao, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.
- [30] J. Peng, Q. Xie, Q. Zhao, Y. Wang, D. Meng, and Y. Leung, "Enhanced 3DTV regularization and its applications on hyper-spectral image denoising and compressed sensing," Sep. 2018, *arXiv:1809.06591*. [Online]. Available: <https://arxiv.org/abs/1809.06591>
- [31] Y. Chen, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Hyperspectral image restoration using framelet-regularized low-rank nonnegative matrix factorization," *Appl. Math. Model.*, vol. 63, pp. 128–147, Nov. 2018.
- [32] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 60–65.
- [33] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [34] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [35] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2862–2869.
- [36] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3952–3966, Sep. 2012.
- [37] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Denoising of multispectral images via nonlocal groupwise spectrum-PCA," in *Proc. Conf. Colour Graph., Imag., Vis.*, Jan. 2010, pp. 261–266.
- [38] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2949–2956.
- [39] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1692–1700.
- [40] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 1791–1798.
- [41] H. Ji, S. Huang, Z. Shen, and Y. Xu, "Robust video restoration by joint sparse and low rank matrix approximation," *SIAM J. Imag. Sci.*, vol. 4, no. 4, pp. 1122–1142, 2011.
- [42] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, Feb. 2013.

- [43] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
- [44] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted Schatten  $p$ -norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.
- [45] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.
- [46] 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.
- [47] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4677–4690, Oct. 2016.
- [48] J.-L. Wang, T.-Z. Huang, T.-H. Ma, X.-L. Zhao, and Y. Chen, "A sheared low-rank model for oblique stripe removal," *Appl. Math. Comput.*, vol. 360, pp. 167–180, Nov. 2019.
- [49] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
- [50] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.
- [51] T. Yokota and H. Hontani, "Simultaneous visual data completion and denoising based on tensor rank and total variation minimization and its primal-dual splitting algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3843–3851.
- [52] T.-Y. Ji, N. Yokoya, X. X. Zhu, and T.-Z. Huang, "Nonlocal tensor completion for multitemporal remotely sensed images' inpainting," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3047–3061, Jun. 2018.
- [53] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.
- [54] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, no. 9, pp. 50–63, Sep. 2013.
- [55] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.
- [56] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3842–3849.
- [57] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2015.
- [58] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.
- [59] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-Y. Ji, and T.-H. Ma, "Tensor N-tubal rank and its convex relaxation for low-rank tensor recovery," Dec. 2018, *arXiv:1812.00688*. [Online]. Available: <https://arxiv.org/abs/1812.00688>
- [60] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.
- [61] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [62] T.-H. Ma, Y. Lou, and T.-Z. Huang, "Truncated  $l_{1-2}$  models for sparse recovery and rank minimization," *SIAM J. Imag. Sci.*, vol. 10, no. 3, pp. 1346–1380, 2017.
- [63] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, and J. Huang, "Hyperspectral image denoising via convex low-fibered-rank regularization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2019, pp. 222–225.
- [64] D. Wei, A. Wang, X. Feng, B. Wang, and B. Wang, "Tensor completion based on triple tubal nuclear norm," *Algorithms*, vol. 11, no. 7, p. 94, Jun. 2018.
- [65] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. IEEE Amer. Control Conf.*, vol. 3, Jun. 2003, pp. 2156–2162.
- [66] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 3441–3473, Nov. 2012.
- [67] Y. Deng, Q. Dai, R. Liu, Z. Zhang, and S. Hu, "Low-rank structure learning via nonconvex heuristic recovery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 3, pp. 383–396, Mar. 2013.
- [68] W. Dong, G. Shi, X. Hu, and Y. Ma, "Nonlocal sparse and low-rank regularization for optical flow estimation," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4527–4538, Oct. 2014.
- [69] T.-Y. Ji, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, and L.-J. Deng, "A non-convex tensor rank approximation for tensor completion," *Appl. Math. Model.*, vol. 48, pp. 410–422, Aug. 2017.



**Yu-Bang Zheng** received the B.S. degree in information and computing science from the Anhui University of Finance and Economics, Bengbu, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

His research interest includes sparse and low-rank modeling for image processing problems. <http://www.esience.cn/people/YuBangZheng/index.html>.



**Ting-Zhu Huang** received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China, in 1986, 1992, and 2001, respectively.

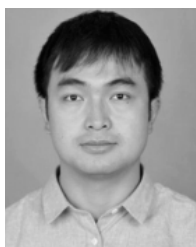
He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Dr. Huang is an Editor of the *Scientific World Journal*, *Advances in Numerical Analysis*, the *Journal of Applied Mathematics*, the *Journal of Pure and Applied Mathematics: Advances in Applied Mathematics*, and the *Journal of Electronic Science and Technology*, China.



**Xi-Le Zhao** received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively.

He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include image processing, computer vision, and machine learning.



**Tai-Xiang Jiang** received the B.S. degree in mathematics and applied mathematics and the Ph.D. degree in mathematics from the University of Electronic Science and Technology of China, Chengdu, China, in 2013 and 2019, respectively.

From 2017 to 2018, he was a co-training Ph.D. student with the Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal, supported by the China Scholarship Council. In 2019, he was a Research Assistant with the Department of Mathematic, Hong Kong Baptist University, Hong Kong, for three months. His research interests include sparse and low-rank modeling, tensor decomposition, and deep learning. <https://sites.google.com/view/taixiangjiang/>.



**Teng-Yu Ji** received the B.S. and Ph.D. degrees from the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China, in 2012 and 2018, respectively.

He is currently an Assistant Professor with the Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, China. His research interests include tensor decomposition and applications, including tensor completion and remotely sensed image reconstruction.



**Tian-Hui Ma** received the B.S. and Ph.D. degrees in applied mathematics from the University of Electronic Science and Technology of China, Chengdu, China, in 2011 and 2017, respectively.

Since 2016, he has been a Visiting Student with The University of Texas at Dallas, Dallas, TX, USA. He is currently a Post-Doctoral Researcher with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China. His research interests include inverse problems in image processing and sparse optimization.