



Fast algorithm with theoretical guarantees for constrained low-tubal-rank tensor recovery in hyperspectral images denoising

Xi-Le Zhao^a, Hao Zhang^{a,*}, Tai-Xiang Jiang^b, Michael K. Ng^c, Xiong-Jun Zhang^d

^a School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

^b FinTech Innovation Center, Financial Intelligence and Financial Engineering Research Key Laboratory of Sichuan province, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

^c Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong

^d School of Mathematics and Statistics and Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan 430079, China

ARTICLE INFO

Article history:

Received 18 October 2019

Revised 16 May 2020

Accepted 5 July 2020

Available online 14 July 2020

Communicated by Wang QI

Keywords:

Tensor tubal rank

Non-convex optimization

Bilateral random tensors projections

Hyperspectral images

ABSTRACT

Hyperspectral images (HSIs) are unavoidably degraded by mixed noise, including Gaussian noise and sparse noise. In this paper, we consider a constrained tubal rank and sparsity model (CTSD) to tackle the HSIs mixed noise removal, which characterizes the clean HSIs via the low-tubal-rank constraint and the sparse noise via the l_0 and l_∞ norm constraints, respectively. Due to the strong non-convexity, the CTSD model is challenging to solve. To tackle the CTSD, we develop the proximal alternating minimization (PAM) algorithm via the exact tensor singular value decomposition (t-SVD) and establish the global convergence under mild assumptions. Since the t-SVD is computationally expensive, especially for large scale images, we further design an efficient inexact PAM algorithm via an inexact t-SVD. The inexact PAM enjoys two advantages: (1) The computational complexity for SVDs of the inexact PAM ($O(m_1 n_2 n_3)$) is about twofold faster than that of the exact PAM ($O(\min(n_1, n_2) n_1 n_2 n_3)$) for $r \ll \min(n_1, n_2)$; (2) The accuracy of the inexact PAM is theoretically guaranteed. Extensive experiments on HSIs denoising demonstrate that the exact and inexact methods both outperform comparing methods in quantitative evaluation metrics and visual effects, and the inexact PAM can compromise between the accuracy and efficiency for large scale HSIs.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Hyperspectral images (HSIs) [1,2] are widely applied in segmentation [3], matching [4], classification [5,6], and spectral signature unmixing [1,7–9]. HSIs often inevitably suffer from various types of noise in the acquisition and transmission process [10,11], which dramatically affects some subsequent applications. Therefore, it is critical to reduce the noise in HSIs and improve their quality for the subsequent image analysis.

Compared with the noise of gray-scale or RGB images, the noise of HSIs is more complex, and there usually exists a combination of different types of noise. As shown in Fig. 1, the corrupted HSI is a 3-D cube, which can be decomposed as the clean HSI part, the sparse noise part, and the Gaussian noise part. The sparse noise part generally includes impulse noise, stripes, and deadlines. Complex noise increases the difficulty of noise removal. Additionally,

since HSIs are usually large scale, HSIs denoising is highly time-consuming. Thus, developing effective and efficient methods for HSIs denoising becomes an important research topic [12–16].

Due to the spatial and spectral redundancy, clean HSIs are usually low-rank. Therefore, many denoising methods are devoted to exploiting the low-rank structure of HSIs [10,11,17]. One possible way is unfolding HSIs to matrices and considering the low-matrix-rank-based methods, such as the classical low-rank matrix recovery (LRMR) [18] using the Godec algorithm [19]. Some non-convex low-matrix-rank regularization were also considered in the low-matrix-rank-based methods to improve denoising performance, including log-sum penalty [20], Schatten p -norm [21], and γ -norm [22]. Additionally, many low-matrix-rank subspace representation-based methods [23–25] were also developed, such as fast hyperspectral image denoising [23,24] and adaptive hyperspectral mixed noise removal [25], which have significant efficiency for large scale HSIs denoising. Although these low-matrix-rank-based methods achieved good effectiveness, unfolding HSIs to matrices destroys the structure of HSIs and introduces the loss of useful structure information. Specifically, unfolding an

* Corresponding author.

E-mail addresses: 201821110218@std.uestc.edu.cn (H. Zhang), mng@maths.hku.hk (M.K. Ng), xjzhang@mail.ccnu.edu.cn (X.-J. Zhang).

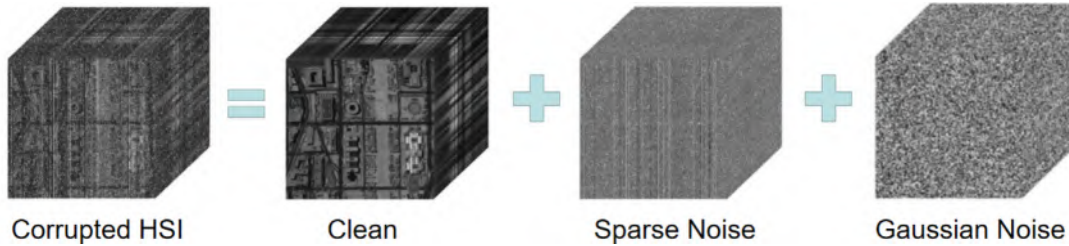


Fig. 1. The decomposition of the corrupted HSI.

HSI along its spectral dimension makes it difficult to utilize the spatial redundancy.

Since tensor can well exploit the global redundancy of the higher-dimensional data, tensor-based methods [26–30] have received considerable attention recently. In the tensor case, the corrupted HSI is a 3-D tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which can be decomposed as

$$\mathcal{X} = \mathcal{L} + \mathcal{S} + \mathcal{N}, \quad (1)$$

where the tensor $\mathcal{L} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the clean HSI part, the tensor $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the sparse noise part, and the tensor $\mathcal{N} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the Gaussian noise part. How to exploit the low-tensor-rank structure of HSIs is the central issue of the low-tensor-rank-based HSIs denoising [31–33]. Different from the matrix, the definition of tensor rank is not unique. Based on different decomposition schemes, including the parallel factor analysis (PARAFAC) decomposition and the Tucker decomposition, several low-tensor-rank-based methods have been developed for HSIs denoising. The PARAFAC decomposition is a natural generation from the definition of matrix rank, where the tensor is decomposed as the sum of the rank-one tensors [34]. The PARAFAC rank [31,34] is defined as the minimal number of the rank-one tensors. However, determining the PARAFAC rank of a given tensor is an NP-hard problem [34]. The Tucker decomposition scheme decomposes a tensor into a core tensor multiplied by matrices along each mode [35]. The Tucker rank is defined as the vector, which consists of the ranks of each unfolding matrix along different modes. Since the Tucker rank is easy to be optimized by matricization along each mode, it inspired many methods for HSIs denoising, including the genetic kernel Tucker decomposition [36], low- n -rank tensor approximation (LRTA) [35], and the Tucker rank and TV minimization [37]. However, the matricization along each mode unavoidably destroys the global intrinsic structure of tensor data.

Recently, the tensor singular value decomposition (t-SVD) [38] introduced a new popular tensor tubal rank, which well preserves the global intrinsic low-rank structure of tensor data [26,39]. Many low-tubal-rank-based methods have been developed for HSI denoising [32,40,41]. Fan et al. (LRTR [32]) considered the tensor nuclear norm (TNN) and the l_1 norm to characterize the clean HSI and the sparse noise, respectively, and employed alternating direction method of multipliers (ADMM) for solving. However, since the TNN and the l_1 norm are convex surrogates of the tensor tubal rank and the l_0 norm [26], they are not the most accurate to characterize the clean HSI and the sparse noise, resulting in a sub-optimal performance in denoising. Further, Fan et al. suggested a TNN and tubal rank hybrid model and solved it by the augmented Lagrange method (ALM) [41]. However, the solving algorithm lacked a rigorous convergence guarantee.

Additionally, many deep-learning-based methods [12–14] are also developed for HSIs denoising. The deep residual convolutional neural network (CNN) was considered in [12] to exploit the spatial and spectral redundancy of HSIs. In [14], Li et al. proposed the unsupervised fully convolutional network, which utilized the

change detection maps to train the deep CNN and then removed the noise during the end-to-end training process. The denoised HSIs can improve the performance of the subsequent image analysis, such as the accuracy of classification [12,41,42]. Wang et al. [42] utilized the locality constraint criterion and the structure-preserving strategy to sufficiently exploit the local and global structures of HSIs and then accurately classified HSIs. Therefore, HSIs denoising is an essential task for many real-world applications.

1.1. Problems and motivations

In this work, we consider the low-tubal-rank-based model for HSIs mixed noise removal, which can be formulated as

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}} \quad & \frac{1}{2} \|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F^2, \\ \text{s.t. } \quad & \mathcal{L} \in \mathcal{C} = \{\mathcal{L} : \text{rank}_t(\mathcal{L}) \leq r\}, \\ & \mathcal{S} \in \mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q\}, \end{aligned} \quad (2)$$

where \mathcal{X} is the observed HSI, \mathcal{L} is the clean HSI, \mathcal{S} is the sparse noise, $\|\cdot\|_F$ is the Frobenius norm, $\text{rank}_t(\cdot)$ is the tensor tubal rank, and $\|\cdot\|_0$ is the l_0 norm. Here, \mathcal{C} is the set of low-tubal-rank tensors \mathcal{L} , where r is the bound of the tubal rank. Here, \mathcal{D} is the set of sparse tensors \mathcal{S} , where q is the bound of the number of non-zero entries.

Since the pixel values of sparse noise are in a given range, we can further introduce the l_∞ norm constraint into the low-tubal-rank-based HSI denoising model, i.e.,

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}} \quad & \|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F^2, \\ \text{s.t. } \quad & \mathcal{L} \in \mathcal{C} = \{\mathcal{L} : \text{rank}_t(\mathcal{L}) \leq r\}, \\ & \mathcal{S} \in \mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}, \end{aligned} \quad (3)$$

where the introduced l_∞ norm $\|\cdot\|_\infty$ is the maximum of absolute values of the tensor's entries. Here, we denote the set of low-tubal-rank tensors as \mathcal{C} , where r is the bound of the tubal rank. We denote the set of sparse tensors as \mathcal{D} , where q is the bound of the number of non-zero entries and c is the bound of the max absolute value of entries. We name the optimization (3) as Constrained Tubal rank and Sparsity tensor Decomposition (CTSD). The CTSD can also flexibly incorporate different regularizers by exploiting desired priors for better performance. Here, we consider the original CTSD in this work.

There are two challenges for solving CTSD:

(1) *Theoretical aspect:* The rigorous convergence is still missing for solving CTSD. Due to the tubal-rank constraint and the l_0 norm constraint, the CTSD is a strongly non-convex problem. ADMM and ALM factually have limitations to solve CTSD from the above discussion. This motivates us to develop the algorithm with theoretical guarantees of convergence for solving CTSD.

(2) *Computational aspect:* Solving CTSD suffers the heavy computational burden for large scale HSIs. Solving the low-tubal-rank tensor of CTSD needs to calculate the exact t-SVD. The exact t-SVD computes the singular value decomposition (SVD) within

each frontal slice of the tensor in the Fourier domain [26]. This motivates us to develop an efficient algorithm to solve the low-tubal-rank tensor of CTSD fast.

1.2. Contributions

The contributions of our work are summarized as follows:

- We consider the CTSD model for HSIs mixed noise removal, which characterizes the clean HSI via the low-tubal-rank constraint and the sparse noise via the l_0 and l_∞ norm constraints, respectively.
- We develop the exact proximal alternating minimization (PAM) algorithm via the exact t-SVD to solve the proposed model and establish the global convergence under mild assumptions.
- Furthermore, we suggest the inexact truncated t-SVD by designing bilateral random tensor projections (t-BRP) and the corresponding inexact PAM for large scale HSIs. The computational complexity for SVDs in the inexact PAM ($O(rn_1n_2n_3)$) is about twofold faster than that of the exact PAM ($O(\min(n_1, n_2)n_1n_2n_3)$) for $r \ll \min(n_1, n_2)$. Moreover, we establish the theoretical guarantee of the accuracy of the inexact PAM.

1.3. Paper Organization

This paper is arranged as follows. Section 2 gives notations and preliminaries. Section 3 discusses the exact PAM and its convergence analysis, and the inexact PAM and its accuracy analysis. Section 4 reports experimental examples, discusses the influence of parameters, and compares the inexact PAM with the exact PAM. Section 5 concludes this paper.

2. Notations and preliminaries

In this section, some notations and preliminaries are introduced for the development of the algorithm for CTSD.

Throughout this paper, we denote tensor as capitalized calligraphic letters (e.g., \mathcal{X}), matrices as capitalized boldface letters (e.g., \mathbf{X}), vectors as boldface lowercase letters (e.g., \mathbf{x}), and scalars as lowercase letters (e.g., x). For a 3-D tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its (i, j, k) -th entry is represented by x_{ijk} , its tube is represented by $\mathcal{A}(i, j, :)$, and its i -th frontal slice is represented by $\mathcal{A}^{(i)}$ or $\mathcal{A}(:, :, i)$. The Frobenius norm of \mathcal{X} is defined as $\|\mathcal{X}\|_F = \sqrt{\sum_{ijk} |x_{ijk}|^2}$ [26]. The l_0 norm of \mathcal{X} is defined as $\|\mathcal{X}\|_0 = \#\{i, j, k : \mathcal{X}(i, j, k) \neq 0\}$. The l_∞ norm of \mathcal{X} is defined as $\|\mathcal{X}\|_\infty = \max_{ijk} |x_{ijk}|$ [26]. The identity tensor can be denoted as $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$.

The related notations of t-SVD and tubal rank are introduced. We denote $\bar{\mathcal{X}}$ as the tensor generated by the Discrete Fourier Transformation (DFT) on each tube of \mathcal{X} , i.e., $\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$. The inverse of $\bar{\mathcal{X}}$ is \mathcal{X} , i.e., $\mathcal{X} = \text{ifft}(\bar{\mathcal{X}}, [], 3)$. We can define the block diagonal matrix \bar{X} of $\bar{\mathcal{X}}$ [26], i.e.,

$$\bar{X} = \text{bdiag}(\bar{\mathcal{X}}) \triangleq \begin{pmatrix} \bar{X}^{(1)} & & & \\ & \bar{X}^{(2)} & & \\ & & \ddots & \\ & & & \bar{X}^{(n_3)} \end{pmatrix}, \tag{4}$$

where the i -th block is the i -th frontal slice of $\bar{\mathcal{X}}$, and we define the block circulant matrix [26] as $\text{bcirc}(\mathcal{X})$, i.e.,

$$\text{bcirc}(\mathcal{X}) \triangleq \begin{pmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \dots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \dots & \mathcal{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \dots & \mathcal{X}^{(1)} \end{pmatrix}. \tag{5}$$

The $\text{bcirc}(\mathcal{X})$ admits the following decomposition, i.e.,

$$(F_{n_3} \otimes I_{n_1}) \cdot \text{bcirc}(\mathcal{X}) \cdot (F_{n_3}^{-1} \otimes I_{n_2}) = \bar{X}, \tag{6}$$

where \otimes is the Kronecker product and F_{n_3} is the DFT matrix.

Now we can define the tensor-tensor product and the corresponding tensor operations as follows:

Definition 1. (Tensor-tensor product [26]) For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times l \times n_3}$, the product of them can be represented as

$$\mathcal{X} * \mathcal{Y} = \text{fold}(\text{bcirc}(\mathcal{X}) \cdot \text{unfold}(\mathcal{Y})), \tag{7}$$

where the new tensor is of the size $n_1 \times l \times n_3$. For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the unfold operator maps \mathcal{X} to the matrix of the size $n_1 n_2 \times n_3$ and fold is the inverse operator of unfold , i.e.,

$$\text{unfold}(\mathcal{X}) = \begin{pmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{pmatrix} \text{ and } \mathcal{X} = \text{fold}(\text{unfold}(\mathcal{X})). \tag{8}$$

Definition 2. (Identity tensor) The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor whose first frontal slice is the $n \times n$ identity matrix, and whose other frontal slices are all zeros.

Definition 3. (Inverse of tensor [26]) We denote $\mathcal{X}^{-1} \in \mathbb{R}^{n \times n \times n_3}$ as the inverse tensor of $\mathcal{X} \in \mathbb{R}^{n \times n \times n_3}$, which satisfies $\mathcal{X}^{-1} * \mathcal{X} = \mathcal{X} * \mathcal{X}^{-1} = \mathcal{I}$.

Definition 4. (Orthogonal tensor [26]) The orthogonal tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfies $\mathcal{X}^H * \mathcal{X} = \mathcal{X} * \mathcal{X}^H = \mathcal{I}$, where \mathcal{X}^H is the conjugate transpose of the orthogonal tensor of the size $n_2 \times n_1 \times n_3$. The conjugate transpose of \mathcal{X} is obtained by conjugate transposing each frontal slice and then reversing the order of transposed frontal slices 2 through, i.e.,

$$\begin{aligned} \mathcal{X}^H(:, :, 1) &= (\mathcal{X}(:, :, 1))^H, \\ \mathcal{X}^H(:, :, n_3 + 2 - i) &= (\mathcal{X}(:, :, i))^H, i = 2, \dots, n_3. \end{aligned} \tag{9}$$

Based on the tensor-tensor product, we can define the t-SVD.

Definition 5. (T-SVD [26]) For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD operation decomposes \mathcal{X} as

$$\mathcal{X} = \mathcal{U} * \mathcal{F} * \mathcal{V}^H, \tag{10}$$

where \mathcal{U} and \mathcal{V} are orthogonal tensors of the size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$, respectively, and \mathcal{F} is an f -diagonal tensor of the size $n_1 \times n_2 \times n_3$ satisfying that all frontal slices are diagonal.

By the t-SVD, we have the tensor multi-rank and tensor tubal rank.

Definition 6. (Tensor multi-rank [26]) The multi-rank of tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector r of the size n_3 , whose i -th element is the rank of the i -th frontal slice of $\bar{\mathcal{X}}$, i.e., $r_i = \text{rank}(\bar{X}^{(i)})$.

Definition 7. (Tensor tubal rank [26]) The tensor tubal rank $\text{rank}_t(\mathcal{X})$ is defined as the number of non-zero singular tubes from \mathcal{F} . That is

$$\text{rank}_t(\mathcal{X}) = \#\{i : \mathcal{F}(i, i, :) \neq 0\} = \max_i r_i. \tag{11}$$

Remark 1. For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have $\text{rank}_t(\mathcal{X}) \leq \min(n_1, n_2)$ and $\text{rank}_t(\mathcal{X} * \mathcal{Y}) \leq \min(\text{rank}_t(\mathcal{X}), \text{rank}_t(\mathcal{Y}))$.

Here, we give the convex surrogate of the multi-rank.

Definition 8. (TNN [26]) The TNN of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be represented as

$$\|\mathcal{X}\|_{\text{TNN}} = \sum_{i=1}^{n_3} \|\bar{X}^{(i)}\|_*, \tag{12}$$

where $\|\cdot\|_*$ is the matrix nuclear norm.

Definition 9. (Tensor spectral norm [26]) The tensor spectral norm of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\|\mathcal{X}\| := \|\text{bcirc}(\mathcal{X})\|$ and we have

$$\|\mathcal{X}\| = \|\text{bcirc}(\mathcal{X})\| = \|\bar{X}\|_2, \tag{13}$$

where $\|\cdot\|_2$ is the matrix spectral norm.

Now, we briefly introduce the necessary PAM framework for solving the CTSD model. PAM is an efficient algorithm framework for solving the non-convex problem with multi-variables [43]. Let the integer $p \geq 2$, and let n_1, \dots, n_p be positive integers. The vector x belongs to the space $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$, denoted by $x = (x_1, \dots, x_p)$, where each $x_i \in \mathbb{R}^{n_i}$. The minimization of functions $f : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p} \rightarrow \mathbb{R} \cup \{+\infty\}$ has the following structure:

$$f(x) = Q(x_1, \dots, x_p) + \sum_{i=1}^p f_i(x_i), \tag{14}$$

where Q is a C^1 function with locally Lipschitz continuous gradient, and $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper lower semi-continuous function, $i = 1, 2, \dots, p$. For each $i = 1, 2, \dots, p$, a bounded sequence of symmetric positive definite matrices $(B_i^t)_{t \in \mathbb{N}}$ of size n_i . Then, the update of x_i is

$$x_i^{t+1} \in \arg \min \left\{ f(x_1^{t+1}, \dots, x_{i-1}^{t+1}, u_i, x_{i+1}^t, \dots, x_p^t) + \frac{1}{2} \langle B_i^t (u_i - x_i^t), u_i - x_i^t \rangle : u_i \in \mathbb{R}^{n_i} \right\}. \tag{15}$$

3. Exact PAM and Inexact PAM for CTSD

In this section, we develop the exact PAM and inexact PAM for solving CTSD and establish corresponding theoretical guarantees.

First, we introduce indicator functions [44] of \mathcal{L} and \mathcal{S} . For closed subsets $\mathcal{C} = \{\mathcal{L} : \text{rank}_t(\mathcal{L}) \leq r\}$ and $\mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}$, we denote $\delta_{\mathcal{C}}$ and $\delta_{\mathcal{D}}$ as indicator functions, i.e.,

$$\delta_{\mathcal{C}}(\mathcal{L}) = \begin{cases} 0 & \text{if } \mathcal{L} \in \mathcal{C} = \{\mathcal{L} : \text{rank}_t(\mathcal{L}) \leq r\}, \\ +\infty & \text{otherwise,} \end{cases} \tag{16}$$

$$\delta_{\mathcal{D}}(\mathcal{S}) = \begin{cases} 0 & \text{if } \mathcal{S} \in \mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}, \\ +\infty & \text{otherwise.} \end{cases}$$

Algorithm 1 Exact PAM for solving CTSD.

Input: $\mathcal{X}, r, q, c, \lambda, \mu$, and ϵ .
Output: \mathcal{L}, \mathcal{S} .
 1: **Initialize** $\mathcal{L}^0 = \mathcal{S}^0 = \mathbf{0}, t = 0$.
 2: **While** not converged **do**
 3: Update \mathcal{L}^{t+1} via (20).
 4: Update \mathcal{S}^{t+1} via (21).
 5: Let $t = t + 1$.
 6: Check the convergence condition:
 7: $\|\mathcal{L}^{t+1} - \mathcal{L}^t\|_F^2 / \|\mathcal{L}^t\|_F^2 \leq \epsilon$.
 8: **End while**

Then, we reformulate (3) as the following unconstrained minimization, i.e.,

$$\min_{\mathcal{L}, \mathcal{S}} \delta_{\mathcal{C}}(\mathcal{L}) + \delta_{\mathcal{D}}(\mathcal{S}) + \frac{1}{2} \|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F^2. \tag{17}$$

By minimizing (17), we have $\delta_{\mathcal{C}}$ and $\delta_{\mathcal{D}}$ to be zero. If $\delta_{\mathcal{C}}$ and $\delta_{\mathcal{D}}$ are zero, \mathcal{L} and \mathcal{S} belong to the constrained sets \mathcal{C} and \mathcal{D} , respectively. Thus, the unconstrained minimization (17) is equal to the CTSD model (3).

3.1. Exact PAM for CTSD and its convergence analysis

Now, we design the exact PAM to solve CTSD and establish the theoretical convergence. Under the PAM framework (15), \mathcal{L} and \mathcal{S} are iteratively updated as follows:

$$\mathcal{L}^{t+1} = \arg \min_{\mathcal{U} \in \mathcal{C}} \left\{ \frac{1}{2} \|\mathcal{X} - \mathcal{U} - \mathcal{S}^t\|_F^2 + \frac{1}{2\lambda} \|\mathcal{U} - \mathcal{L}^t\|_F^2 \right\}, \tag{18}$$

$$\mathcal{S}^{t+1} = \arg \min_{\mathcal{V} \in \mathcal{D}} \left\{ \frac{1}{2} \|\mathcal{X} - \mathcal{L}^{t+1} - \mathcal{V}\|_F^2 + \frac{1}{2\mu} \|\mathcal{V} - \mathcal{S}^t\|_F^2 \right\}. \tag{19}$$

Thus, we obtain the following sub-problems and alternately solve them until convergence, i.e.,

$$\mathcal{L}^{t+1} = P_{\mathcal{C}} \left(\frac{\lambda^{-1} \mathcal{L}^t - \mathcal{S}^t + \mathcal{X}}{\mathcal{I} + \lambda^{-1} \mathcal{I}} \right), \tag{20}$$

$$\mathcal{S}^{t+1} = P_{\mathcal{D}} \left(\frac{\mu^{-1} \mathcal{S}^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1} \mathcal{I}} \right), \tag{21}$$

where $P_{\mathcal{C}}$ and $P_{\mathcal{D}}$ are projections onto $\mathcal{C} = \{\mathcal{L} : \text{rank}_t(\mathcal{L}) \leq r\}$ and $\mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}$ [44], respectively. We summarize the exact PAM for CTSD in Algorithm 1.

For \mathcal{L} subproblem (20), we solve it by computing the first- r truncated t-SVD of $(\lambda^{-1} \mathcal{L}^t - \mathcal{S}^t + \mathcal{X}) / (\mathcal{I} + \lambda^{-1} \mathcal{I})$. The following theorem guarantees the best tubal rank- r approximation.

Theorem 1. [45] For a given tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, if $\mathcal{X}_r = \arg \min_{\tilde{\mathcal{X}} \in \mathcal{C}} \|\mathcal{X} - \tilde{\mathcal{X}}\|_F$ and $\mathcal{C} = \{\tilde{\mathcal{X}} : \text{rank}_t(\tilde{\mathcal{X}}) \leq r\}$, the theoretical optimal error $\|\mathcal{X} - \mathcal{X}_r\|_F$ can be given as

$$\|\mathcal{X} - \mathcal{X}_r\|_F = \left(\frac{1}{n_3} \sum_{i=1}^{n_3} \sum_{j>r} (\hat{\sigma}_j^{(i)})^2 \right)^{1/2}, \tag{22}$$

where $\mathcal{X}_r = \mathcal{U}_r * \mathcal{F}_r * \mathcal{V}_r^H$ is the first- r truncated t-SVD of \mathcal{X} and $\bar{\sigma}_j^{(i)} = \bar{\mathcal{F}}(j, j, i)$ is the j -th singular value in i -th frontal slice in the Fourier domain.

For \mathcal{S} subproblem (21), we solved it by the entry-wise hard thresholding operator and projection onto $\mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}$. Specifically, for $\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}$, the entry-wise hard thresholding operator $\mathcal{H}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right)$ sets all but the largest q elements of $|\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}|$ to zero and the (i, j, k) -th entry of the tensor from $\mathcal{H}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right)$ can be denoted as $\mathcal{H}_{ijk}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right)$. Thus, the projection onto $\mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}$ is solved as

$$S_{ijk}^{t+1} = \begin{cases} c & \mathcal{H}_{ijk}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right) > c, \\ \mathcal{H}_{ijk}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right) & |\mathcal{H}_{ijk}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right)| \leq c, \\ -c & \mathcal{H}_{ijk}^q\left(\frac{\mu^{-1}S^t - \mathcal{L}^{t+1} + \mathcal{X}}{\mathcal{I} + \mu^{-1}\mathcal{I}}\right) < -c. \end{cases} \quad (23)$$

• The convergence analysis of exact PAM for CTSD

Now, denoting $\delta_{\mathcal{L}}(\mathcal{L}) + \delta_{\mathcal{D}}(\mathcal{S}) + \frac{1}{2}\|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F^2$ by $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}^t, \mathcal{S}^t)$ and $\frac{1}{2}\|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F^2$ by $N(\mathcal{L}, \mathcal{S})$, we are ready to establish the theoretical guarantee of the convergence for the exact PAM.

Theorem 2. The sequence $(\mathcal{L}^t, \mathcal{S}^t)$ generated by Algorithm 1 converges to a critical point of $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}^t, \mathcal{S}^t)$.

For the proof of Theorem 2, the following conditions are required [43]:

- $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ has the K – \mathbb{L} property at each $(\mathcal{L}^t, \mathcal{S}^t)$.
- $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ is a proper lower semi-continuous function.
- \mathcal{L} and \mathcal{S} are bounded and satisfy the sufficient decrease condition and relative error condition.

Proof. The three conditions are dividedly proven as follows.

Firstly, we show that $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ has the K – \mathbb{L} property at each $(\mathcal{L}^t, \mathcal{S}^t)$ by proving that $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ is a semi-algebraic function. Since the multi-rank is defined in the Fourier domain actually, we reformulate the CTSD in the Fourier domain, i.e.,

$$\begin{aligned} \min_{\bar{\mathcal{L}}, \bar{\mathcal{S}}} \|\bar{\mathcal{X}} - \bar{\mathcal{L}} - \bar{\mathcal{S}}\|_F^2, \text{ s.t. } \bar{\mathcal{L}} \in \bar{\mathcal{C}} = \{\bar{\mathcal{L}} : \text{rank}(\bar{\mathcal{L}}^{(i)}) \leq r\}, \\ \bar{\mathcal{S}} \in \bar{\mathcal{D}} = \left\{ \bar{\mathcal{S}} : \left\| \begin{pmatrix} \mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_1} \end{pmatrix} \bar{\mathcal{S}} \begin{pmatrix} \mathbf{F}_{n_3} \otimes \mathbf{I}_{n_2} \end{pmatrix} \right\|_0 \right. \\ \left. \leq q, \left\| \begin{pmatrix} \mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_1} \end{pmatrix} \bar{\mathcal{S}} \begin{pmatrix} \mathbf{F}_{n_3} \otimes \mathbf{I}_{n_2} \end{pmatrix} \right\|_\infty \leq c \right\}, \end{aligned} \quad (24)$$

where $\text{rank}(\cdot)$ is the rank of a matrix. The subset $\bar{\mathcal{C}}$ guarantees that each element of the multi-rank of $\bar{\mathcal{L}}$ is less than r . Since $\bar{\mathcal{S}} = \begin{pmatrix} \mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_1} \end{pmatrix} \bar{\mathcal{S}} \begin{pmatrix} \mathbf{F}_{n_3} \otimes \mathbf{I}_{n_2} \end{pmatrix}$, the subset $\bar{\mathcal{D}}$ is equal to the subset $\mathcal{D} = \{\mathcal{S} : \|\mathcal{S}\|_0 \leq q, \|\mathcal{S}\|_\infty \leq c\}$. We denote $\delta_{\bar{\mathcal{C}}}(\bar{\mathcal{L}})$ and $\delta_{\bar{\mathcal{D}}}(\bar{\mathcal{S}})$ as the indicator functions of $\bar{\mathcal{C}}$ and $\bar{\mathcal{D}}$, i.e.,

$$\begin{aligned} \delta_{\bar{\mathcal{C}}}(\bar{\mathcal{L}}) &= \begin{cases} 0 & \text{if } \bar{\mathcal{L}} \in \bar{\mathcal{C}}, \\ +\infty & \text{otherwise,} \end{cases} \\ \delta_{\bar{\mathcal{D}}}(\bar{\mathcal{S}}) &= \begin{cases} 0 & \text{if } \bar{\mathcal{S}} \in \bar{\mathcal{D}}, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \quad (25)$$

Then, we consider the equivalent unconstrained optimization of the CSTD model in the Fourier domain, i.e.,

$$\min_{\bar{\mathcal{L}}, \bar{\mathcal{S}}} \delta_{\bar{\mathcal{C}}}(\bar{\mathcal{L}}) + \delta_{\bar{\mathcal{D}}}(\bar{\mathcal{S}}) + \frac{1}{2}\|\bar{\mathcal{X}} - \bar{\mathcal{L}} - \bar{\mathcal{S}}\|_F^2. \quad (26)$$

Since the low-rank matrices set is semi-algebraic, the subset $\bar{\mathcal{C}}$ is semi-algebraic. Constraint sets of the l_0 norm and the l_∞ norm are semi-algebraic [46]. The intersection set of semi-algebraic sets is semi-algebraic [46]. Indicator functions of semi-algebraic sets are semi-algebraic functions [46], and the Frobenius norm is a semi-algebraic function [46]. Thus, $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ is a semi-algebraic function. Since the semi-algebraic real-valued function is the K – \mathbb{L} function [46], $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ has the K – \mathbb{L} property at each $(\mathcal{L}^t, \mathcal{S}^t)$.

Secondly, we show that $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ is a proper lower semi-continuous function. It can verify that $N(\mathcal{L}, \mathcal{S})$ is a \mathcal{C}^1 function with locally Lipschitz continuous gradient, and $\delta_{\bar{\mathcal{C}}}(\bar{\mathcal{L}})$ and $\delta_{\bar{\mathcal{D}}}(\bar{\mathcal{S}})$ are proper lower semi-continuous. Therefore, $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$ is a proper lower semi-continuous function.

Thirdly, we show that \mathcal{L} and \mathcal{S} are bounded and satisfy the sufficient decrease condition and relative error condition. We have $\|\mathcal{X}\|_F - \|\mathcal{L}\|_F - \|\mathcal{S}\|_F \leq \|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F$ and $\|\mathcal{L}\|_F \leq \|\mathcal{X} - \mathcal{L} - \mathcal{S}\|_F + \|\mathcal{X}\|_F - \|\mathcal{S}\|_F$ by the triangle inequality. Since the objective function $M_{\mathcal{C},\mathcal{D}}$ monotonically decreases and $\|\mathcal{S}\|_F$ is bounded by $\|\mathcal{S}\|_\infty \leq c$, $\|\mathcal{L}\|_F$ is bounded. Thus, \mathcal{L} and \mathcal{S} are bounded. The process of Algorithm 1 is a special instance of the PAM framework (15) when $\mathbf{B}_i = \lambda \mathbf{I} = \mu \mathbf{I}$ ($\lambda, \mu > 0$) from Remark 6.1 of [43]. Thus, \mathcal{L} and \mathcal{S} generated by Algorithm 1 satisfy the sufficient decrease condition and relative error condition.

In summary, the sequence $(\mathcal{L}^t, \mathcal{S}^t)$ converges to a critical point of $M_{\mathcal{C},\mathcal{D}}(\mathcal{L}, \mathcal{S})$. \square

Algorithm 2 Inexact t-SVD.

Input: $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and Randomized tensor $\mathcal{A}_1 \in \mathbb{R}^{n_2 \times r \times n_3}$.

Output: \mathcal{X}_{BRP}

1: $\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, \cdot, 3)$; $\bar{\mathcal{A}}_1 = \text{fft}(\mathcal{A}_1, \cdot, 3)$;

2: **for** $i = 1 : n_3$ **do**

3: $\bar{\mathbf{Y}}_1^{(i)} = \bar{\mathcal{X}}^{(i)} \bar{\mathcal{A}}_1^{(i)}$, $\bar{\mathcal{A}}_2^{(i)} = \bar{\mathbf{Y}}_1^{(i)}$, $\bar{\mathbf{Y}}_2^{(i)} = \bar{\mathcal{X}}^{H(i)} \bar{\mathcal{A}}_2^{(i)}$, $\bar{\mathbf{Y}}^{(i)} = \bar{\mathcal{X}}^{(i)} \bar{\mathbf{Y}}_2^{(i)}$;

4: $\bar{\mathcal{X}}_{\text{BRP}}(\cdot, \cdot, i) = \bar{\mathcal{X}}_{\text{BRP}}^{(i)} = \bar{\mathbf{Y}}^{(i)} \left[\bar{\mathcal{A}}_2^{H(i)} \bar{\mathbf{Y}}_1^{(i)} \right]^{-1} \bar{\mathbf{Y}}_2^{H(i)}$;

5: **end for**

6: $\mathcal{X}_{\text{BRP}} = \text{ifft}(\bar{\mathcal{X}}_{\text{BRP}}, \cdot, 3)$.

3.2. Inexact PAM for CTSD and its accuracy analysis

Since the t-SVD is computationally expensive especially for large HSIs, we further develop an efficient inexact PAM algorithm and establish the accuracy guarantee. Based on the tensor-tensor product, we suggest the inexact t-SVD to obtain the tubal rank- r approximation for (20) by designing the t-BRP.

Definition 10. (T-BRP) For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ ($n_1 > n_2$), the t-BRP of \mathcal{X} can be constructed, i.e., $\mathcal{Y}_1 = \mathcal{X} * \mathcal{A}_1$ and $\mathcal{Y}_2 = \mathcal{X}^H * \mathcal{A}_2$, wherein $\mathcal{A}_1 \in \mathbb{R}^{n_2 \times r \times n_3}$ and $\mathcal{A}_2 \in \mathbb{R}^{n_1 \times r \times n_3}$ are random tensors.

The tubal rank- r approximation of \mathcal{X} can be calculated as

$$\mathcal{X}_{\text{BRP}} = \mathcal{Y}_1 * \left(\mathcal{A}_2^H * \mathcal{Y}_1 \right)^{-1} * \mathcal{Y}_2^H, \quad (27)$$

where \mathcal{X}_{BRP} is the solution of (20) by the inexact t-SVD and the parameter r directly governs the tubal rank of \mathcal{X}_{BRP} by Remark 1. For the accuracy of the inexact t-SVD, we consider \mathcal{A}_1 as the standard Gaussian tensor, denote the $\mathcal{Y}_1 = \mathcal{X} * \mathcal{A}_1$ as \mathcal{A}_2 , update $\mathcal{Y}_2 = \mathcal{X}^H * \mathcal{A}_2$, and replace the left \mathcal{Y}_1 of (27) by $\mathcal{X} * \mathcal{Y}_2$, i.e.,

$$\mathcal{X}_{\text{BRP}} = \mathcal{X} * \mathcal{X}^H * \mathcal{X} * \mathcal{A}_1 * \left(\mathcal{A}_1^H * \mathcal{X}^H * \mathcal{X} * \mathcal{A}_1 \right)^{-1} * \mathcal{A}_1^H * \mathcal{X}^H * \mathcal{X}. \tag{28}$$

The inexact t-SVD can be efficiently calculated by (6) and (7). For $i = 1, \dots, n_3$, let $\bar{A}_1^{(i)} \in \mathbb{C}^{n_2 \times r}$ is the standard Gaussian matrix [47,48]. The rank- r approximation of $\bar{X}^{(i)}$ is obtained by

$$\bar{X}_{\text{BRP}}^{(i)} = \bar{X}^{(i)} \bar{X}^{H(i)} \bar{X}^{(i)} \bar{A}_1^{(i)} \left(\bar{A}_1^{H(i)} \bar{X}^{H(i)} \bar{X}^{(i)} \bar{A}_1^{(i)} \right)^{-1} \bar{A}_1^{H(i)} \bar{X}^{H(i)} \bar{X}. \tag{29}$$

Thus, we have that

$$\bar{X}_{\text{BRP}} = \bar{X} \bar{X}^H \bar{X} \bar{A}_1 \left(\bar{A}_1^H \bar{X}^H \bar{X} \bar{A}_1 \right)^{-1} \bar{A}_1^H \bar{X}^H \bar{X}. \tag{30}$$

Then, we have $\mathcal{X}_{\text{BRP}} = \left(\mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_1} \right) \bar{X}_{\text{BRP}} \left(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_2} \right)$. The inexact t-SVD is summarized as Algorithm 2.

By the construction of the inexact t-SVD, we can observe that the computational complexity of the inexact t-SVD ($O(rn_1n_2n_3)$) is lower than that of the exact t-SVD ($O(\min(n_1, n_2)n_1n_2n_3)$) in the Fourier domain for $r \ll \min(n_1, n_2)$. Now, we denote the exact PAM-based method as CTSD, which utilizes the exact t-SVD to solve (20), and we denote the inexact PAM-based method as R-CTSD, which solves (20) via the inexact t-SVD. Table 1 gives the computational complexity comparison of different algorithms, including CTSD, R-CTSD, and other comparing methods.

Since the estimated $(\lambda^{-1}\mathcal{L}^t - \mathcal{S}^t + \mathcal{X}) / (\mathcal{I} + \lambda^{-1}\mathcal{I})$ at each iteration may not be accurate enough, it is not necessary to calculate the exact solution of the subproblem (20). For computational efficiency, we can consider the accelerated Algorithm 2.

• **The accuracy analysis of inexact PAM for CTSD**

Now, we establish the accuracy guarantee of the inexact PAM. The error in the inexact PAM mainly arises from the subproblem (20) at each iteration. Thus, we need to discuss the error between the inexact solution and the exact solution of the subproblem (20).

We firstly have the t-SVD of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as

$$\mathcal{X} = \mathcal{U} * \mathcal{F} * \mathcal{V}^H = \mathcal{U}_r * \mathcal{F}_r * \mathcal{V}_r^H + \mathcal{U}_p * \mathcal{F}_p * \mathcal{V}_p^H. \tag{31}$$

We denote the exact solution of (20) generated by first- r truncated t-SVD as $\mathcal{X}_r = \mathcal{U}_r * \mathcal{F}_r * \mathcal{V}_r^H$ and the inexact solution of (20) generated by the inexact t-SVD as \mathcal{X}_{BRP} .

Theorem 3. For a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and a standard Gaussian tensor $\mathcal{A}_1 \in \mathbb{R}^{n_2 \times r \times n_3}$, the error between the inexact solution \mathcal{X}_{BRP} and the exact solution \mathcal{X}_r is theoretically bounded by

$$\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\|^2 \leq \|\bar{F}_p^2 \left(\bar{V}_p^H \bar{A}_1 \right) \left(\bar{V}_p^H \bar{A}_1 \right)^\dagger \bar{F}_p^{-1}\|_2^2, \tag{32}$$

where $\|\cdot\|$ is the tensor spectral norm (Def. 8) and $\|\cdot\|_2$ is the matrix spectral norm.

Remark 2. Since the singular values of each frontal slice $\bar{X}^{(i)}$ decay fast for the clean HSI, the singular values except the first r -th singular values are tiny numbers. Thus, the error bound is very small, and the tubal rank- r approximation by the inexact t-SVD is sufficiently accurate.

For the proof of Theorem 3, we need to introduce some lemmas and propositions as follows:

Lemma 1. [49]. Suppose that $\bar{G} \succeq 0$. For every A , the matrix $A^T G A \succeq 0$. Especially,

$$G \preceq B \Rightarrow A^T G A \preceq A^T B A. \tag{33}$$

Table 1
Computational complexity of different methods.

Method	Computational complexity ($r \ll \min\{n_1, n_2, n_3\}$)
LRMR [18]	$h^2 n_3 r$
LRMA [21]	$\min\{h^2 n_3^2, h^4 B\} + J n_3 + 2h^2 n_3$
LRTR [32]	$n_1 n_2 n_3 \log n_3 + \min(n_1, n_2) n_1 n_2 n_3$
CTSD	$n_1 n_2 n_3 \log n_3 + \min(n_1, n_2) n_1 n_2 n_3$
R-CTSD	$n_1 n_2 n_3 \log n_3 + r n_1 n_2 n_3$

Proposition 1. [49]. Suppose that $\text{range}(B) \subset \text{range}(G)$. For each A , it holds that $\|\mathcal{P}_B A\|_2 \leq \|\mathcal{P}_G A\|_2$ and $\|(I - \mathcal{P}_B)A\|_2 \leq \|(I - \mathcal{P}_G)A\|_2$.

Proposition 2. [49]. Suppose that $\bar{G} \succeq 0$. Thus,

$$I - (I + G)^{-1} \preceq G. \tag{34}$$

Proposition 3. [49]. There is $\|G\|_2 \leq \|A\|_2 + \|C\|_2$ for each partitioned positive semi-definite matrix

$$G = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}. \tag{35}$$

Now, we give the proof of Theorem 3.

Proof. We consider the theoretical error between the inexact solution \mathcal{X}_{BRP} and the exact solution \mathcal{X}_r of the subproblem (20) as the following tensor spectral norm, i.e.,

$$\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\| = \|\mathcal{U}_r * \mathcal{F}_r * \mathcal{V}_r^H - \mathcal{Y}_1 * \left(\mathcal{A}_2^H * \mathcal{Y}_1 \right)^{-1} * \mathcal{Y}_2^H\|. \tag{36}$$

From the tensor spectral norm (Def. 8), we have that

$$\begin{aligned} \|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\| &= \|\text{bcirc}(\mathcal{X}_r - \mathcal{X}_{\text{BRP}})\| = \|\bar{X}_r - \bar{X}_{\text{BRP}}\|_2 \\ &= \|\bar{U}_r \bar{F}_r \bar{V}_r^H - \bar{Y}_1 \left(\bar{A}_2^H \bar{Y}_1 \right)^{-1} \bar{Y}_2^H\|_2 \\ &= \|\bar{U}_r \bar{F}_r \bar{V}_r^H - \bar{X} \bar{A}_1 \left(\bar{A}_2^H \bar{X} \bar{A}_1 \right)^{-1} \bar{A}_2^H \bar{X}\|_2. \end{aligned} \tag{37}$$

The unitary invariance of the spectral norm leads to

$$\begin{aligned} \|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\| &= \|\bar{X}_r - \bar{X}_{\text{BRP}}\|_2 \\ &= \|\bar{U}_r^H \left[\bar{U}_r \bar{F}_r \bar{V}_r^H - \bar{X} \bar{A}_1 \left(\bar{A}_2^H \bar{X} \bar{A}_1 \right)^{-1} \bar{A}_2^H \bar{X} \right] \bar{V}\|_2 \\ &= \|\left(\bar{F}_r \ 0 \right) - \left(\bar{F}_r \ 0 \right) \bar{V}^H \bar{A}_1 \left(\bar{A}_2^H \bar{X} \bar{A}_1 \right)^{-1} \bar{A}_2^H \bar{U} \bar{F}\|_2 \\ &= \left\| \begin{pmatrix} \bar{F}_r & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \bar{F}_r & 0 \\ 0 & 0 \end{pmatrix} \bar{V}^H \bar{A}_1 \left(\bar{A}_2^H \bar{X} \bar{A}_1 \right)^{-1} \bar{A}_2^H \bar{U} \bar{F} \right\|_2 \\ &= \|\bar{F}^* \left[I - \bar{V}^H \bar{A}_1 \left(\bar{A}_2^H \bar{X} \bar{A}_1 \right)^{-1} \bar{A}_2^H \bar{U} \bar{F} \right]\|_2, \end{aligned} \tag{38}$$

where

$$\bar{F}^* = \begin{pmatrix} \bar{F}_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{39}$$

From (28), we have $\mathcal{A}_2 = \mathcal{Y}_1 = \mathcal{X} * \mathcal{A}_1 = \mathcal{U} * \mathcal{F} * \mathcal{V}^H * \mathcal{A}_1$ and $\mathcal{A}_1 = \mathcal{Y}_2 = \mathcal{X}^H * \mathcal{A}_2 = \mathcal{X}^H * \mathcal{X} * \mathcal{A}_1 = \mathcal{U} * \mathcal{F} * \mathcal{F} * \mathcal{V}^H * \mathcal{A}_1$. Thus, $\bar{A}_2 = \bar{U} \bar{F} \bar{V}^H \bar{A}_1$ and $\bar{A}_1 = \bar{V} \bar{F}^2 \bar{V}^H \bar{A}_1$. Then, we have that

$$\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\| = \|\bar{F}^* \left[I - \bar{F}^2 \bar{V}^H \bar{A}_1 \left(\bar{A}_1^H \bar{V} \bar{F}^4 \bar{V} \bar{A}_1 \right)^{-1} \bar{A}_1^H \bar{V} \bar{F}^2 \right]\|_2. \tag{40}$$

We define the orthogonal projector of a given G as $\mathcal{P}_G = G \left(G^H G \right)^{-1} G^H$, which projects a given matrix to the column space (range) of G . Then, we have

$$\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\| = \|\bar{F}^* (I - \mathcal{P}_G)\|_2, G = \bar{F}^2 \bar{V}^H \bar{A}_1. \tag{41}$$

Since $\text{range}(G(\bar{V}_r^H \bar{A}_1)^\dagger \bar{F}_r^{-2}) \subset \text{range}(G)$, we have the following inequality from [Proposition 1](#) ($\|(I - \mathcal{P}_B)A\|_2 \leq \|(I - \mathcal{P}_C)A\|_2$), i.e.,

$$\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\| = \|\bar{F}^*(I - \mathcal{P}_C)\|_2 \leq \|\bar{F}^*(I - \mathcal{P}_B)\|_2, \quad (42)$$

where

$$B = \begin{pmatrix} \bar{F}_r^2 \bar{V}_r^H \bar{A}_1 \\ \bar{F}_p^2 \bar{V}_p^H \bar{A}_1 \end{pmatrix} (\bar{V}_r^H \bar{A}_1)^\dagger \bar{F}_r^{-2} = \begin{pmatrix} I \\ M \end{pmatrix}. \quad (43)$$

We rewrite $I - \mathcal{P}_B$ as

$$I - \mathcal{P}_B = \begin{pmatrix} I - (I + M^H M)^{-1} & -(I + M^H M)^{-1} M^H \\ -M(I + M^H M)^{-1} & I - M(I + M^H M)^{-1} M^H \end{pmatrix}. \quad (44)$$

On the one hand, the top-left block of (44) has $I - (I + M^H M)^{-1} \preceq M^H M$ from [Proposition 2](#) ($I - (I + G)^{-1} \preceq G$). On the other hand, the bottom-right block of (44) has $I - M(I + M^H M)^{-1} M^H \preceq I$ from [Lemma 1](#) ($G \preceq B \Rightarrow A^T G A \preceq A^T B A$). Together, we have

$$I - \mathcal{P}_B = \begin{pmatrix} M^H M & -(I + M^H M)^{-1} M^H \\ -M(I + M^H M)^{-1} & I \end{pmatrix}. \quad (45)$$

Then, we have the following formula by [Lemma 1](#), i.e.,

$$\bar{F}^*(I - \mathcal{P}_B)\bar{F}^* = \begin{pmatrix} \bar{F}_r^H M^H M \bar{F}_r & -\bar{F}_r^H (I + M^H M)^{-1} M^H 0 \\ -0 M (I + M^H M)^{-1} & 0 \end{pmatrix}. \quad (46)$$

By applying [Proposition 3](#), we have that (46) is bounded by

$$\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\|^2 = \|\bar{F}^*(I - \mathcal{P}_B)\bar{F}^*\| \leq \|\bar{F}_p^2 (\bar{V}_p^H \bar{A}_1) (\bar{V}_r^H \bar{A}_1)^\dagger \bar{F}_r^{-1}\|_2^2. \quad (47)$$

The proof is completed. \square

Remark 3. The accuracy of randomized algorithms receives considered attention [50–52]. Zhou et al. [52] discussed the error between the matrix X and its matrix rank- r approximation, i.e., $\|X - X_{\text{BRP}}\|$. In this work, we are interested in the error between the exact solution via the exact t-SVD and the inexact solution via the inexact t-SVD for subproblem (20), i.e., $\|\mathcal{X}_r - \mathcal{X}_{\text{BRP}}\|$, instead of the tensor \mathcal{X} and its tubal rank- r approximation.

4. Experiments

To verify the effectiveness and efficiency of the proposed methods, comprehensive experiments of the synthetic and real data are conducted. We denote the exact PAM-based method as CTSD and the inexact PAM-based method as R-CTSD. The compared methods consist of LRMA [35], BM4D [53], LRMR [18], LRMA [21], and LRTR [32]. LRMA is a low-Tucker-rank-based method, which performs both spatial and spectral low-rank approximation. BM4D considers the spatial-spectral self-similarity and searches similar 3-D cubes to remove noise collaboratively. LRMR is a well-established matrix-based method for mixed noise removal by the Godec algorithm [19]. LRMA is a low-matrix-rank-based method via the non-convex weighted Schatten p -norm. LRTR characterizes the clean HSI via the TNN and the sparse noise via the l_1 norm. We select the parameters of the compared methods as suggested in the reference papers.

Parameters settings of CTSD and R-CTSD: We discuss the parameters settings of CTSD and R-CTSD in the synthetic and real data experiments. For the CTSD and the R-CTSD, the parameters settings are the same. Parameters λ and μ are the weight of

proximal term, which can be determined from the candidate set $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. The parameter c is fixed to be 1 to constrain the pixel values of the sparse noise in the interval $[-1, 1]$. The parameters r and q govern the tubal-rank of the targeted HSI and the sparsity of the sparse noise part, respectively. The specific settings of r and q are listed in [Table 2](#).

All the tests are implemented on the Windows 10 system and Matlab (2017b) with the CPU Intel Core i7-8700 k 3.70 GHz and 16 GB RAM. All the methods are run once for all experiments.

4.1. Synthetic data

In the simulated experiments, two HSIs data were tested, including Washington DC Mall¹ of the size $256 \times 256 \times 191$ and Pavia University¹ of the size $610 \times 340 \times 103$. The two HSIs were band-wisely normalized to be in the range $[0, 1]$. Three metrics served as the quantitative evaluation, including the peak signal-to-noise ratio (PSNR) [54], the structural similarity index (SSIM) [55] in the range $[0, 1]$, and the spectral angle mapping (SAM) [56] in the range $[0, 1]$ (degree). The mean PSNR (MPSNR) and mean SSIM (MSSIM) are generated by averaging these values of PSNR and SSIM from all bands, respectively. Generally, better denoising performances are reflected by higher MPSNR, MSSIM, and lower SAM values.

We simulated the additional noise as the following three cases:

Case 1: Synthetic data was with Gaussian noise, fixed impulse noise, and different stripes and deadlines noise. The zero-mean Gaussian noise was added to all bands with different variances, which randomly distributed between $[0.02, 0.04]$. The impulse noise affecting 20% of pixels was added to each band. Stripes and deadlines were added to selected 11 bands, and the width randomly ranged from one line to three lines.

Case 2: Synthetic data was with fixed Gaussian noise and different impulse noise. The zero-mean Gaussian noise with the fixed variance as 0.02 was added to all bands. The impulse noise was added to each band, and the percentages of pixels affected by impulse noise were respectively set to 10%, 20%, and 30%.

Case 3: Synthetic data was with different Gaussian noise and fixed impulse noise. The zero-mean Gaussian noise was added to all bands, and the variances of Gaussian noise were respectively set to 0.02, 0.04, and 0.06. The impulse noise was added to each band, and the percentage of impulse noise was fixed at 20%.

Quantitative Evaluation: [Tables 3–5](#) list the quantitative comparisons of competing methods in different cases. Specifically, our methods significantly outperform other competing methods for all the evaluation metrics in most cases, e.g., in [Table 3](#), CTSD and R-CTSD achieve around 4.5 dB and 4 dB gain in MPSNR beyond the third-best method (LRMA) on Pavia University. In [Table 4](#), when the impulse noise affects 30% pixels of Washington DC Mall, the proposed methods still achieve around 40 dB in MPSNR, while LRMR, LRMA, and LRTR achieve around 35 dB in MPSNR. As Gaussian noise increases to 0.06 in [Table 5](#), the proposed methods rank the first and the second place in terms of the MPSNR and SAM. Comparing the running time in [Tables 3–5](#), we can observe that R-CTSD is the fastest. Although CTSD is the third-fast in running time and slower than BM4D, it achieves the best quantitative metrics in most cases. Thus, our methods are both of high efficiency in mixed noise removal.

Moreover, we compare the quantitative performance of CTSD and R-CTSD. We can observe that the performance of R-CTSD is close to that of CTSD for all the evaluation metrics. In [Table 5](#), when Gaussian noise is 0.04 and 0.06, R-CTSD even exceeds CTSD in terms of MSSIM. Thus, the accuracy of R-CTSD is sufficient for HSIs

¹ <http://lesun.weebly.com/hyperspectral-data-set.html>.

Table 2

The parameters setting in CTSD and R-CTSD for the synthetic and real data experiments.

Experiments	Case	Data	r	q
Synthetic	Case1	Washington DC Mall Pavia University	5	2.7×10^6 4.0×10^6
	Case2	Washington DC Mall	5	1.5×10^6 (Sparsenoise : 10%) 2.5×10^6 (Sparsenoise : 20%) 3.5×10^6 (Sparsenoise : 30%)
	Case3	Washington DC Mall	5	2.5×10^6
Real	–	HYDICE Urban	3	1.5×10^6
		Hyperion Australia	5	1.5×10^6

Table 3The quantitative evaluation of competing methods in case 1. The **best** values and the **second-best** values are respectively highlighted by bold fronts and underlines.

Data	Metrics	Noisy	LRTA [35]	BM4D [53]	LRMR [18]	LRMA [21]	LRTR [32]	CTSD	R-CTSD
Washington DC Mall	MPSNR	11.32	21.25	22.68	35.42	37.55	36.89	41.06	<u>40.61</u>
	MSSIM	0.118	0.577	0.508	0.950	0.970	0.959	<u>0.984</u>	0.989
	SAM	47.47	15.39	13.69	5.70	4.05	5.03	<u>3.98</u>	2.37
	Time (s)	–	<u>54.12</u>	166.21	289.11	5974.56	147.43	91.25	36.65
Pavia University	MPSNR	11.40	21.87	24.31	33.71	35.27	34.71	39.72	<u>39.29</u>
	MSSIM	0.076	0.453	0.536	0.915	0.927	0.912	0.971	<u>0.969</u>
	SAM	48.69	15.45	11.18	6.13	5.24	7.27	<u>2.91</u>	2.94
	Time (s)	–	<u>76.51</u>	207.13	597.32	13451.67	274.61	153.24	58.89

Table 4The quantitative evaluation of competing methods in case 2. The **best** values and the **second-best** values are respectively highlighted by bold fronts and underlines.

Data	Sparse Noise	Metrics	Noisy	LRTA [35]	BM4D [53]	LRMR [18]	LRMA [21]	LRTR [32]	CTSD	R-CTSD
Washington DC Mall	10%	MPSNR	14.35	22.55	25.43	39.51	40.06	40.25	42.84	<u>41.97</u>
		MSSIM	0.252	0.535	0.616	<u>0.986</u>	0.982	0.980	0.999	0.999
		SAM	41.69	18.61	11.32	3.48	3.32	3.25	2.30	<u>2.33</u>
	20%	MPSNR	11.36	20.30	22.78	35.25	37.59	38.80	41.34	<u>40.83</u>
		MSSIM	0.120	0.464	0.517	0.966	0.975	0.973	0.981	<u>0.980</u>
		SAM	47.36	18.11	14.51	6.52	3.42	3.76	3.27	<u>3.42</u>
	30%	MPSNR	9.61	18.98	19.38	34.34	35.72	35.93	40.75	<u>39.86</u>
		MSSIM	0.071	0.532	0.403	0.927	0.944	0.940	<u>0.970</u>	0.972
		SAM	49.67	15.50	17.16	6.95	5.17	5.52	<u>4.38</u>	4.25
	Average time (s)	–	<u>52.39</u>	168.35	285.21	5928.76	148.01	87.08	33.65	

Table 5The quantitative evaluation of competing methods in case 3. The **best** values and the **second-best** values are respectively highlighted by bold fronts and underlines.

Data	Gaussian noise	Metrics	Noisy	LRTA [35]	BM4D [53]	LRMR [18]	LRMA [21]	LRTR [32]	CTSD	R-CTSD
Washington DC Mall	0.02	MPSNR	11.36	20.30	22.78	35.25	37.59	38.80	40.42	<u>39.60</u>
		MSSIM	0.120	0.464	0.517	0.966	0.975	0.973	0.981	<u>0.980</u>
		SAM	47.36	18.11	14.51	6.52	<u>3.42</u>	3.76	3.27	<u>3.42</u>
	0.04	MPSNR	11.31	20.40	22.70	35.14	35.67	35.80	37.82	<u>37.05</u>
		MSSIM	0.117	0.476	0.512	0.947	0.951	<u>0.957</u>	0.956	0.963
		SAM	47.51	17.69	14.64	5.80	5.16	4.70	4.49	<u>4.54</u>
	0.06	MPSNR	11.22	21.45	22.56	32.19	34.70	33.04	35.02	<u>34.86</u>
		MSSIM	0.112	0.606	0.507	0.915	0.928	0.921	<u>0.927</u>	0.929
		SAM	47.72	14.16	14.85	7.83	6.62	6.40	6.31	<u>6.52</u>
	Average time (s)	–	<u>54.47</u>	173.47	263.17	6108.35	141.32	135.27	33.96	

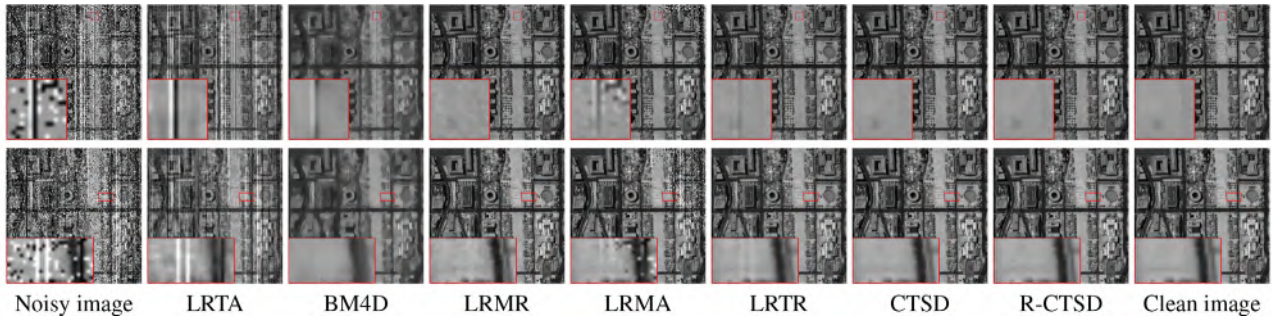


Fig. 2. The denoising images by different methods for Washington DC Mall with mixed noise in case 1. The first row is the result in band 77 of Washington DC Mall and the second row is the result in band 79 of Washington DC Mall.

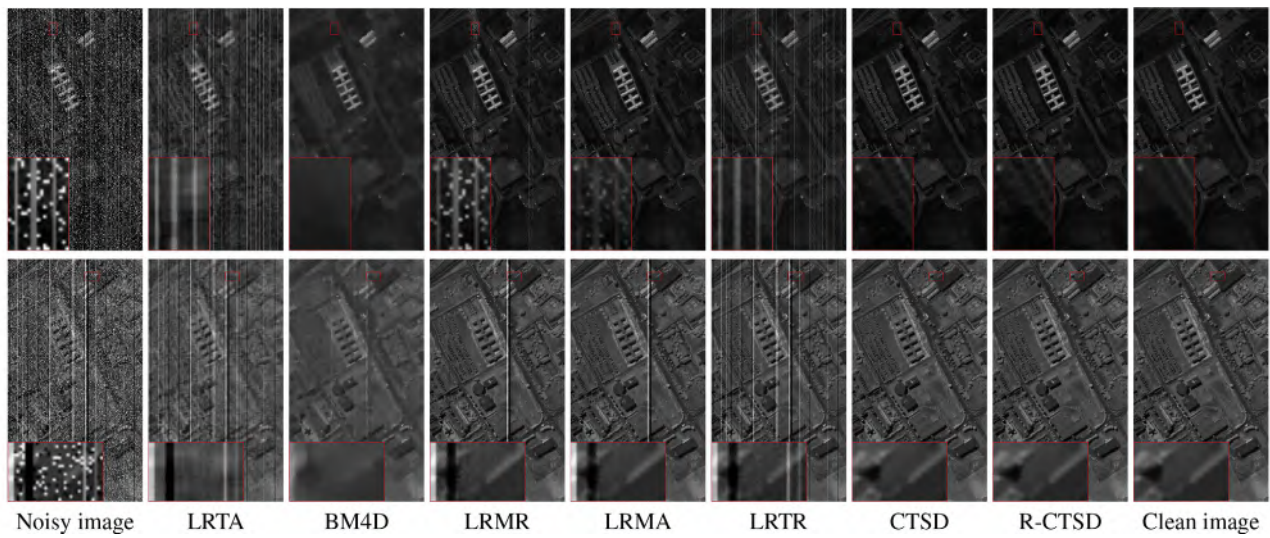


Fig. 3. The denoising images by different methods for Pavia University with mixed noise in case 1. The first row is the results in band 10 of Pavia University and the second row is the results in band 79 of Pavia University.

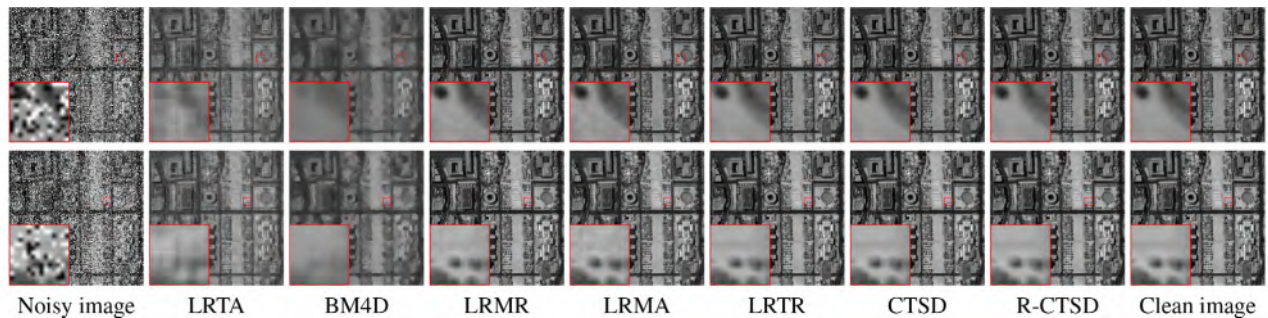


Fig. 4. The denoising images by different methods for Washington DC Mall with 30% sparse noise in case 2. The first row is the results in band 70 of Washington DC Mall and the second row is the results in band 82 of Washington DC Mall.

denoising. Additionally, for the running time in Tables 3–5, R-CTSD is about twofold faster than CTSD. Therefore, R-CTSD compromises between accuracy and efficiency.

Visual Quality Evaluation: Figs. 2–5 show the denoising images by different methods. Specifically, Figs. 2 and 3 show some denoising bands of Washington DC Mall and Pavia University in case 1. We can observe that the proposed CTSD and R-CTSD remove the mixed noise completely and preserve the details well. The denoising results of compared methods can be summarized as follows, i.e., remaining apparent noise and losing some details. For instance, BM4D removes almost all of the noise but blurs the images in

Figs. 2 and 3. Notably, there exist apparent stripes and deadlines in the results of LRMR, LRMA, and LRTR in Fig. 3. The following possible reasons maybe cause this noise to remain. For LRMR and LRMA, since stripes and deadlines could simultaneously be sparse and low-rank, matrix-based methods can not separate the low-rank part and the sparse noise part well. For LRTR, TNN is the convex surrogate of the tensor tubal rank, resulting in the sub-optimal performance in vision. In contrast, our methods directly optimize the tubal rank and separate the clean HSI and noise well. Thus, CTSD and R-CTSD significantly outperform compared methods visually.

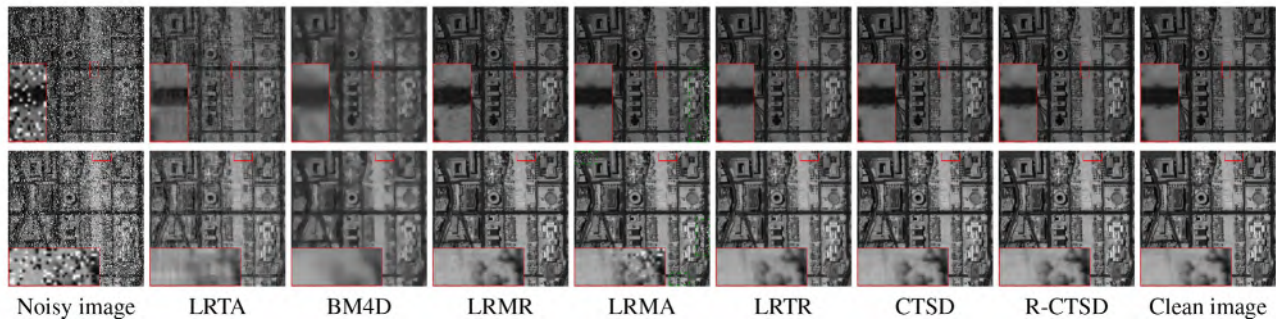


Fig. 5. The denoising images by different methods for Washington DC Mall with 0.04 noise variance in case 3. The first row is the results in band 76 of Washington DC Mall and the second row is the results in band 77 of Washington DC Mall.

Figs. 4 and 5 show some denoising bands of Washington DC Mall in case 2 and 3. In summary, the compared methods remove almost all of the noise. Zooming in an area of denoising images, we can observe that LRTA and BM4D remove noise but cause details missing. In Figs. 4 and 5, LRMR, LRMA, LRTR, and our methods achieve good visual performance in most denoising images, except that LRMA contains little sparse noise in Fig. 5. This indicates that these low-rank-based methods can effectively remove Gaussian noise with the existence of impulse noise.

4.2. Real data

Two real corrupted HSIs are considered for the real data experiment, including HYDICE urban data² and Earth Observing(EO)-1 Hyperion L1R-level Australia data.³

HYDICE urban data: The original size of this real data is $304 \times 304 \times 210$. Fig. 6. shows the denoising results of bands 103 and 206, which are severely corrupted by Gaussian noise and stripes. We can observe that BM4D, LRMR, LRMA, LRTR, and the proposed methods obtain similar visual results and remove almost all of the noise, while the result of LRTA contains much Gaussian noise. From the zoomed-in area, we can find that the result of BM4D loses some details, and the results of LRMR, LRMA, LRTR still contain a small number of stripes. It is easy to see that our methods remove stripes completely and preserve the details well.

EO-1 Hyperion Australia data: The original size of this real data is $256 \times 400 \times 150$. Fig. 7 shows the denoising results of bands 44 and 48. From the zoomed-in area, we can observe that our methods remove the noise well while compared methods remain visible stripes or Gaussian noise. The results of BM4D contain fewer stripes than those of LRMR, LRMA, and LRTR, but some details are blurred. The proposed CTSD and R-CTSD effectively remove the mixed noise and finely preserve the details.

4.3. Discussion

The influence of parameters r and q : The parameters r and q govern the solution of the CTSD model. The value r directly constrains the tubal rank of the clean HSI, and q denotes the sparsity of the sparse noise part, i.e., the number of pixels contaminated by impulse noise, stripes, and deadlines. Fig. 8 displays the influence of the parameters r and q in different cases using the inexact PAM. We can observe that the performance is affected by the parameters r and q .

The comparison between CTSD and R-CTSD: Finally, we compare the performance of the CTSD and the R-CTSD. Fig. 9 displays the curves of the running time, the relative error of the decomposition [19], and the PSNR at each iteration from CTSD and R-CTSD. Specifically, the relative error at each iteration is defined as

$$\|\mathcal{X} - \mathcal{L}^t - \mathcal{S}^t\|_F / \|\mathcal{X}\|_F. \quad (48)$$

In Fig. 9, as the iteration increases, the relative error and PSNR of the CTSD are better than those of R-CTSD, because CTSD obtains the low-tubal-rank tensor by the exact t-SVD. In contrast, R-CTSD obtains the low-tubal-rank tensor via the inexact t-SVD. Since the computational complexity of the inexact t-SVD is lower than that of the exact t-SVD, the running time of R-CTSD is about two-fold faster than that of CTSD in Fig. 9. Therefore, R-CTSD compromises between accuracy and efficiency.

5. Conclusion

In this paper, we considered the CTSD model for HSIs denoising. Due to the tubal-rank constraint and the l_0 and l_∞ norm constraints, the CTSD model was challenging to solve. To tackle the CTSD model, we developed the exact PAM algorithm via the exact t-SVD and established the theoretical convergence. We further suggested an efficient inexact PAM algorithm via designing an inexact t-SVD for large HSIs and establish its accuracy guarantee. The comprehensive experimental examples of the synthetic and real data demonstrated that both exact PAM and inexact PAM achieved excellent performance on HSIs noise removal. Moreover, the quantitative evaluation metrics and visual effects of the inexact PAM were close to those of the exact PAM, while the running time of the inexact PAM was significantly less than that of the exact PAM. Thus, the inexact PAM can compromise between accuracy and efficiency.

CRediT authorship contribution statement

Xi-Le Zhao: Methodology, Investigation, Writing - original draft. **Hao Zhang:** Data curation, Software, Visualization, Writing - original draft. **Tai-Xiang Jiang:** Conceptualization, Writing - review & editing, Methodology. **Michael K. Ng:** Supervision, Writing - review & editing. **Xiong-Jun Zhang:** Writing - review & editing, Investigation, Validation.

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.

² <http://www.tec.army.mil/hypercube>.

³ <http://remote-sensing.nci.org.au/>.

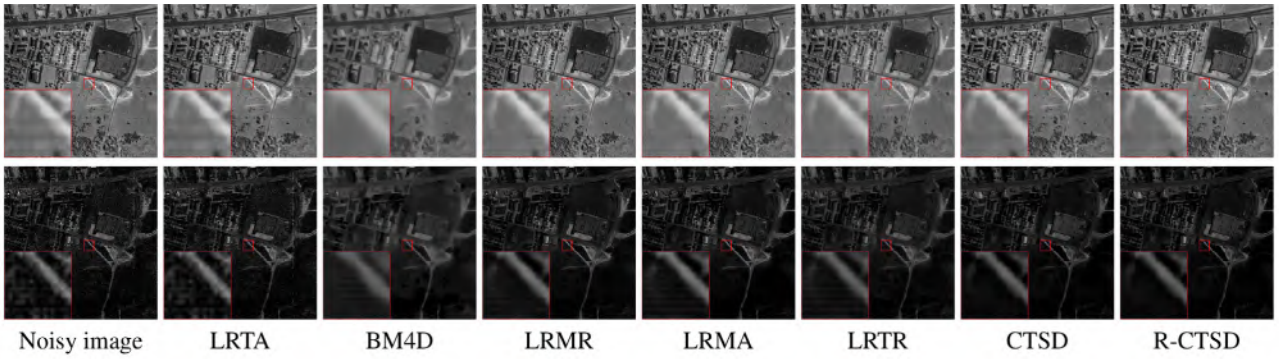


Fig. 6. The denoising images by different methods for HYDICE urban. The first row is the results in band 103 of HYDICE urban and the second row is the results in band 206 of HYDICE urban.

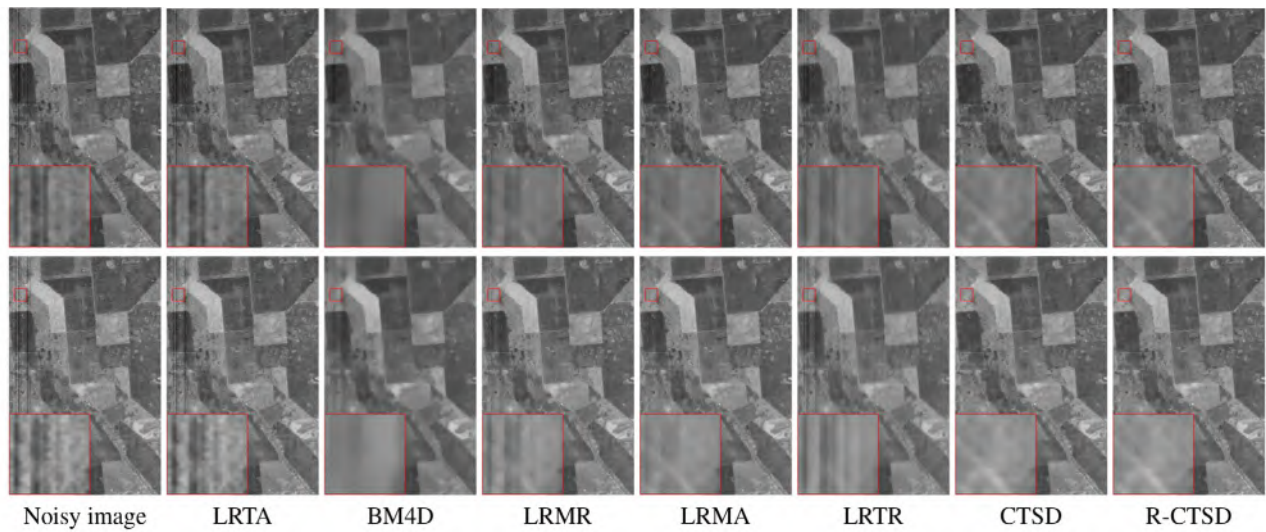


Fig. 7. The denoising images by different methods for EO-1 Hyperion Australia. The first row is the results in band 44 of EO-1 Hyperion Australia and the second row is the results in band 48 of EO-1 Hyperion Australia.

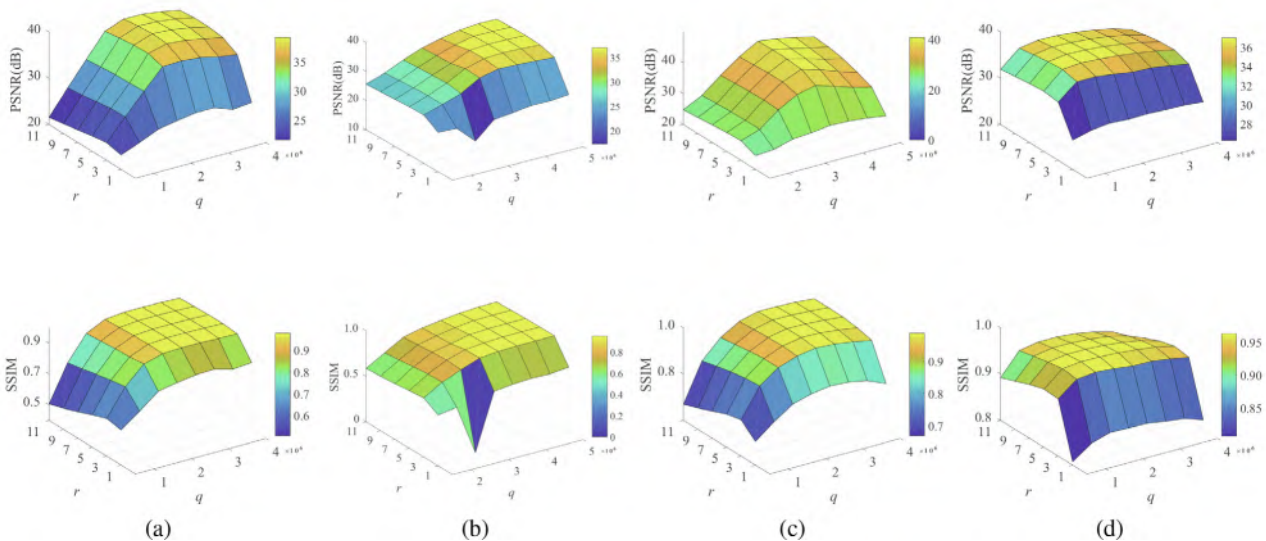


Fig. 8. Recovery results under different values of the parameters r and q . (a) is the change of the PSNR and SSIM from case 1 in the Washington DC Mall; (b) is the change of the PSNR and SSIM from case 1 in the Pavia University; (c) is the change of the PSNR and SSIM from case 2 with 30% sparse noise; (d) is the change of the PSNR and SSIM from case 3 with 0.04 Gaussian noise.

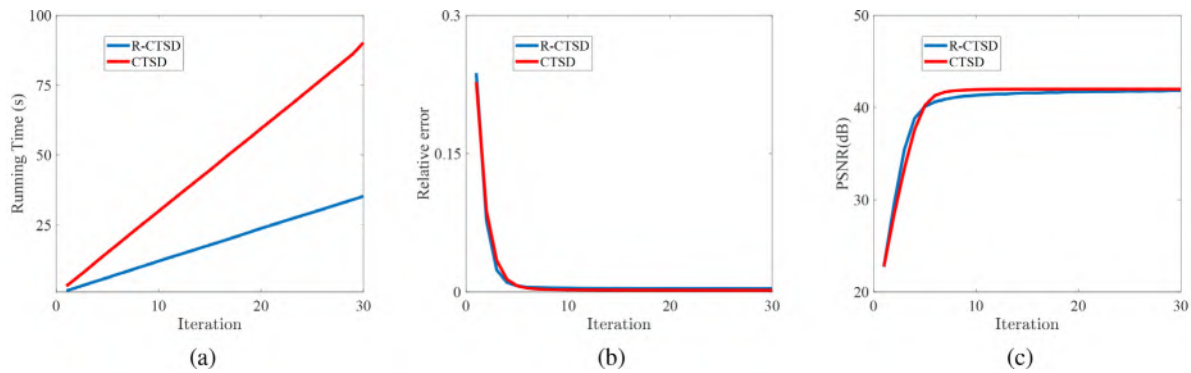


Fig. 9. The curves of running time, the relative error, and the PSNR of the CTSD and the R-CTSD with the change of iteration for Washington DC Mall in case 1. (a) is the change of the running time (s); (b) is the change of the relative error ($\|\mathcal{X} - \mathcal{S}^t - \mathcal{L}^t\|_F / \|\mathcal{X}\|_F$); (c) is the change of the PSNR (dB).

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61876203, 61772003, 11801206, and 11871025), HKRGC GRF (12306616, 12200317, 12300519, and 12300218), HKU Grant (104005583), Hubei Provincial Natural Science Foundation of China under grant (2018CFB105), and Fundamental Research Funds for the Central Universities under grant (CCNU19ZN017).

References

- [1] J.M. Bioucas-Dias, A. Plaza, J. Chanussot, Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 5 (2) (2012) 354–379.
- [2] J.M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, Hyperspectral remote sensing data analysis and future challenges, *IEEE Geosci. Remote Sens. Mag.* 1 (2) (2013) 6–36.
- [3] Z. Zhang, E. Pasolli, M.M. Crawford, J.C. Tilton, An active learning framework for hyperspectral image classification using hierarchical segmentation, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 9 (2) (2016) 640–654.
- [4] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, J. Tian, Robust feature matching for remote sensing image registration via locally linear transforming, *IEEE Trans. Geosci. Remote Sens.* 53 (12) (2015) 6469–6481.
- [5] H. Zhang, J. Li, Y. Huang, L. Zhang, A nonlocal weighted joint sparse representation classification method for hyperspectral imagery, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 7 (6) (2014) 2056–2065.
- [6] H. Zhang, H. Zhai, L. Zhang, P. Li, Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 54 (6) (2016) 3672–3684.
- [7] X. Lu, H. Wu, Y. Yuan, P. Yan, X. Li, Manifold regularized sparse NMF for hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 51 (5) (2013) 2815–2826.
- [8] H. We, H. Zhang, L. Zhang, Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 9 (9) (2016) 4267–4279.
- [9] B. Zhang, L. Zhuang, L. Gao, W. Luo, Q. Ran, Q. Du, PSO-EM: A hyperspectral unmixing algorithm based on normal compositional model, *IEEE Trans. Geosci. Remote Sens.* 52 (12) (2014) 7782–7792.
- [10] Q. Yuan, L. Zhang, H. Shen, Hyperspectral image denoising employing a spectral-spatial adaptive total variation model, *IEEE Trans. Geosci. Remote Sens.* 50 (10) (2012) 3660–3677.
- [11] W. He, H. Zhang, L. Zhang, H. Shen, Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration, *IEEE Trans. Geosci. Remote Sens.* 54 (1) (2016) 178–188.
- [12] Q. Yuan, Q. Zhang, J. Li, H. Shen, L. Zhang, Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network, *IEEE Trans. Geosci. Remote Sens.* 57 (2) (2019) 1205–1218.
- [13] Y. Chang, L. Yan, H. Fang, S. Zhong, W. Liao, HSI-DeNet: Hyperspectral image restoration via convolutional neural network, *IEEE Trans. Geosci. Remote Sens.* 57 (2) (2019) 667–682.
- [14] X. Li, Z. Yuan, Q. Wang, Unsupervised deep noise modeling for hyperspectral image change detection, *Remote Sens.* 11 (3) (2019) 258.
- [15] 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.
- [16] 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.* (2020), <https://doi.org/10.1109/TGRS.2020.2987954>.
- [17] W. He, H. Zhang, L. Zhang, H. Shen, Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 8 (6) (2015) 3050–3061.
- [18] H. Zhang, W. He, L. Zhang, H. Shen, Q. Yuan, Hyperspectral image restoration using low-rank matrix recovery, *IEEE Trans. Geosci. Remote Sens.* 52 (8) (2014) 4729–4743.
- [19] T. Zhou, D. Tao, Godec: Randomized low-rank and sparse matrix decomposition in noisy case, in: *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 33–40.
- [20] E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted L_1 minimization, *J. Fourier Anal. Appl.* 14 (5) (2007) 877–905.
- [21] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, W. Zhang, Hyperspectral image restoration via iteratively regularized weighted Schatten p -norm minimization, *IEEE Trans. Geosci. Remote Sens.* 54 (8) (2016) 4642–4659.
- [22] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, G. He, Denoising of hyperspectral images using nonconvex low rank matrix approximation, *IEEE Trans. Geosci. Remote Sens.* 55 (9) (2017) 5366–5380.
- [23] L. Zhuang, J.M. Bioucas-Dias, Fast hyperspectral image denoising based on low rank and sparse representations, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 11 (3) (2018) 730–742.
- [24] L. Zhuang, J.M. Bioucas-Dias, Fast hyperspectral image denoising based on low rank and sparse representations. In *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)* (2016) 3847, 1850.
- [25] T.-X. Jiang, L. Zhuang, T.-Z. Huang, J.M. Bioucas-Dias, Adaptive hyperspectral mixed noise removal, in: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2018, pp. 4035–4038.
- [26] 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.* 42 (4) (2020) 925–938.
- [27] 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.* 400 (2020) 137–149.
- [28] W. Cao, Y. Wang, C. Yang, X. Chang, Z. Han, Z. Xu, Folded-concave penalization approaches to tensor completion, *Neurocomputing* 152 (25) (2015) 261–273.
- [29] 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), <https://doi.org/10.1016/j.amc.2019.124783>, 124783.
- [30] 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.* 372 (2020), <https://doi.org/10.1016/j.cam.2019.112680>, 112680.
- [31] X. Guo, X. Huang, L. Zhang, L. Zhang, Hyperspectral image noise reduction based on rank-1 tensor decomposition, *ISPRS J. Photogr. Remote Sens.* 83 (1) (2013) 50–63.
- [32] H. Fan, Y. Chen, Y. Guo, H. Zhang, G. Kuang, Hyperspectral image restoration using low-rank tensor recovery, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 10 (10) (2017) 4589–4604.
- [33] 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.
- [34] X. Liu, S. Bourennane, C. Fossati, Denoising of hyperspectral images using the parafac model and statistical performance analysis, *IEEE Trans. Geosci. Remote Sens.* 50 (10) (2012) 3717–3724.
- [35] N. Renard, S. Bourennane, J. Blanc-Talon, Denoising and dimensionality reduction using multilinear tools for hyperspectral images, *IEEE Geosci. Remote Sens. Lett.* 5 (2) (2008) 138–142.
- [36] A. Karami, M. Yazdi, A.Z. Asli, Noise reduction of hyperspectral images using kernel non-negative Tucker decomposition, *IEEE J. Sel. Topics Signal Process.* 46 (7) (2011) 487–493.
- [37] T. Yokota, 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 Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3843–3851.
- [38] Z. Zhang, G. Ely, S. Aeron, N. Hao, M. Kilmer, Novel methods for multilinear data completion and denoising based on tensor-svd, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 3842–3849.

- [39] Z. Zhang, S. Aeron, Exact tensor completion using t-svd, *IEEE Trans. Signal Process.* 65 (6) (2017) 1511–1526.
- [40] H. Zhang, X.-L. Zhao, T.-X. Jiang, M.K. Ng, Constrained low-tubal-rank tensor recovery for hyperspectral images mixed noise removal by bilateral random projections, in: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2019, pp. 1939–1942.
- [41] H. Fan, C. Li, Y. Guo, G. Kuang, J. Ma, Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising, *IEEE Trans. Geosci. Remote Sens.* 56 (10) (2018) 6196–6213.
- [42] Q. Wang, X. He, X. Li, Locality and structure regularized low rank representation for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 57 (2) (2019) 911–923.
- [43] H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.* 137 (1) (2013) 91–129.
- [44] H. Attouch, B. Jerome, P. Redont, A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality, *Math. Oper. Res.* 35 (2) (2010) 438–457.
- [45] M. Kilmer, C. Martin, Factorization strategies for third-order tensors, *Linear Algebra Appl.* 435 (3) (2013) 148–172.
- [46] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.* 46 (1) (2014) 459–494.
- [47] M. Fazel, E. Candes, B. Recht, P. Parrilo, Compressed sensing and robust recovery of low rank matrices, in: *Proc. 42nd Asilomar Conf. Signals Syst. Comput (ACSSC)*, 2008, pp. 1043–1047.
- [48] F. Woolfe, E. Liberty, V. Rokhlin, M. Tygert, A fast randomized algorithm for the approximation of matrices, *Appl. Comput. Harmon. Anal.* 25 (3) (2008) 335–366.
- [49] N. Halko, P.G. Martinsson, J.A. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.* 53 (2) (2010) 217–288.
- [50] M. Che, Y. Wei, Randomized algorithms for the approximations of tucker and the tensor train decompositions, *Adv. Comput. Math.* 45 (1) (2019) 395–428.
- [51] C. Battaglino, G. Ballard, G.K. Tamara, A practical randomized cp tensor decomposition, *SIAM J. Math. Anal. Appl.* 39 (2) (2017) 876–901.
- [52] T. Zhou, D. Tao, Bilateral random projections, in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2012, pp. 1286–1290.
- [53] M. Maggioni, V. Katkovnik, K. Egiazarian, A. Foi, Nonlocal transform-domain filter for volumetric data denoising and reconstruction, *IEEE Trans. Image Process.* 22 (1) (2013) 119–133.
- [54] A. Horé, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, 2010, pp. 23–26.
- [55] 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.
- [56] R.H. Yuhas, J.W. Boardman, A.F. Goetz, Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques, in: *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, 1993, pp. 205–208.



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. He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests are model-driven and data-driven methods for image processing problems.



Hao Zhang received the B.S. degrees in the college of science from the Sichuan Agricultural University (SAU), Ya'an, China, in 2018. He is currently pursuing the M.S. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China. His research interests are modeling and algorithm for high-order data recovery based on low rank prior of tensors.



Tai-Xiang Jiang received the B.S. and Ph.D. degrees in mathematics and applied mathematics from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013. He is currently a lecturer with the School of Economic Information Engineering, Southwestern University of Finance and Economics. His research interests include sparse and low-rank modeling, tensor decomposition and multi-dimensional image processing.



Michael K. Ng is the Director of Research Division for Mathematical and Statistical Science, and Chair Professor of Department of Mathematics, the University of Hong Kong. His research areas are data science, scientific computing, and numerical linear algebra.



Xiong-Jun Zhang received the Ph.D. degree from the College of Mathematics and Econometrics, Hunan University, Changsha, China, in 2017. From 2015 to 2016, he was an Exchange Ph.D. Student with the Department of Mathematics, Hong Kong Baptist University, Hong Kong. He is currently an Assistant Professor with the School of Mathematics and Statistics, Central China Normal University, Wuhan, China. His current research interests include image processing and tensor optimization.