Nonlocal Tensor-Ring Decomposition for Hyperspectral Image Denoising

Yong Chen, Wei He, Member, IEEE, Naoto Yokoya, Member, IEEE, Ting-Zhu Huang, and Xi-Le Zhao

Abstract—Hyperspectral image (HSI) denoising is a fundamental problem in remote sensing and image processing. Recently, nonlocal low-rank tensor approximation-based denoising methods have attracted much attention due to their advantage of being capable of fully exploiting the nonlocal self-similarity and global spectral correlation. Existing nonlocal low-rank tensor approximation methods were mainly based on two common decomposition [Tucker or CANDECOMP/PARAFAC (CP)] methods and achieved the state-of-the-art results, but they are subject to certain issues and do not produce the best approximation for a tensor. For example, the number of parameters for Tucker decomposition increases exponentially according to its dimensions, and CP decomposition cannot better preserve the intrinsic correlation of the HSI. In this article, a novel nonlocal tensor-ring (TR) approximation is proposed for HSI denoising by using TR decomposition to explore the nonlocal self-similarity and global spectral correlation simultaneously. TR decomposition approximates a high-order tensor as a sequence of cyclically contracted third-order tensors, which has strong ability to explore these two intrinsic priors and to improve the HSI denoising results. Moreover, an efficient proximal alternating minimization algorithm is developed to optimize the proposed TR decomposition model efficiently. Extensive experiments on three simulated data sets under several noise levels and two real data sets verify that the proposed TR model provides better HSI denoising results than several state-of-the-art methods in terms of quantitative and visual performance evaluations.

Index Terms—Denoising, hyperspectral image (HSI), nonlocal self-similarity, tensor-ring (TR) decomposition.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) is acquired by imaging spectroscopy over hundreds of bands. This provides richer spectral information than color images, which is conductive to the fine description of real scenes. However, due to the sensor instability, photon effects, calibration error, and physical mechanism, an HSI is unavoidably contaminated by noise during the imaging process [1]. The noise in the HSI not only destroys the image vision but also influences the succeeding application tasks, such as image fusion [2], [3] and unmixing [4]. Therefore, noise removal from the HSI is a vital topic in the areas of remote sensing and image processing.

Recently, many HSI denoising approaches have been proposed, and they can be approximately grouped into three categories: 2-D image denoising-based methods, 3-D tensor image denoising-based methods, and nonlocal tensor approximation denoising-based methods. The 2-D image denoising method is a classical approach in image processing. The related 2-D approaches, such as total variation (TV) methods [5], [6] and sparse representation [7], can be directly extended to process the HSI separately band by band. These methods can effectively remove low-level noise. However, when the noise is strong, these methods cannot achieve satisfactory denoising results. To remove the different intensities of noise better, the 2-D extended methods were proposed that group similar local patches in the image and achieve better results. These methods include nonlocal sparse representation and low-rank matrix approximation methods [8]–[12]. However, they fail to consider the intrinsic characteristics of global spectral correlation in the HSI, resulting in the spectral distortion of a denoised image. By considering the spectral correlation in the HSI, many researchers have employed low-rank regularization to the HSI by reshaping and stacking all images into a matrix [13]–[19]. Moreover, to avoid computation of singular value decomposition and improve the efficiency, low-rank matrix factorization was also applied to spectral correlation depiction for HSI noise removal [20]–[24]. These methods can effectively preserve the spectral information in the denoising process but fail to preserve the spatial structure.

An HSI is a stack of several 2-D images, and it contains three dimensions: one spectral dimension and two spatial dimensions. Thus, it can be modeled as three-order tensor data. Naturally, tensor-based models can improve the denoising results, because they can fully capture the spatial–spectral correlation of the HSI. There are three kinds of tensor decompositions employed for image denoising: Tucker decomposition [25]–[29],CANDECOMP/PARAFAC (CP) decomposition [1], and tensor singular value decomposition [30]. The advantages of these three tensor decomposition models are that they can simultaneously explore the spatial–spectral correlation between the HSIs within all the bands and better...
Fig. 1. Framework of the proposed HSI denoising method.

maintain the spatial–spectral structure in the image. However, the mentioned tensor-based methods ignored the nonlocal self-similarity prior in HSI denoising [31], which is an important prior for improving the denoising performance.

Nonlocal self-similarity indicates the fact that there are many repeated local patterns across the HSI, and the strong correlation exists not only in spectral dimension but also in the nonlocal dimension by grouping the similar patches. The nonlocal similar patch group processing can significantly help the reconstruction of the degraded image [10]. To take full advantage of the nonlocal self-similarity prior in HSI denoising, many nonlocal-based HSI denoising methods have been proposed [31]–[36]. As shown in Fig. 1, nonlocal-based methods try to segment the image into overlapping patches and cluster the similar patches into groups. In [37], the similar nonlocal groups are denoised by a Wiener filter. Peng et al. [31] first proposed a tensor dictionary learning (TDL) method to process nonlocal similar groups, which could fully capture the nonlocal self-similarity and global spectral low-rank priors. On the basis of the TDL architecture, many state-of-the-art nonlocal tensor-related approaches were proposed for HSI denoising, including the intrinsic tensor sparsity (ITS) measure method [32], [35] and the hyper-Laplacian regularized unidirectional low-rank tensor recovery problem (LLRT) [33]. With the ITS method, it was thought that the nonlocal group tensors should obtain the property of Tucker and CP decompositions, and a core tensor sparsity-regularized Tucker decomposition model was proposed. For the LLRT method, it was suggested that the correlation of the nonlocal grouped tensor in the spatial and spectral dimensions was not strong enough and only employed the weighted nuclear norm minimization [10] to explore the correlation property along with the nonlocal mode. Although these two methods achieved good denoising results, the ITS regularization is complex with too many regularizations, while LLRT fails to explore the global spectral information of an HSI. He et al. [34] illustrated that the denoising performance can be improved by balancing the correlations between the spectral and nonlocal self-similarities. Thus, there is still much room to improve the performance of HSI denoising.

Nonlocal-based methods are focused on the correlation representation of the grouped three-order tensor, namely local spatial, global spectral, and nonlocal self-similarity correlations. Utilization and balancing of the correlations between the three modes is an important problem. The ITS method is based on the Tucker decomposition and the CP decomposition. Tucker decomposition represents a tensor using one core tensor and a set of factor matrices, and it can capture the correlation of three modes by minimizing the Tucker rank. However, the number of parameters (variables) increases exponentially following its dimensions. Assume that an n-order tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n}$ and the Tucker rank is $[R, \cdots, R]$; then, the number of variables of the Tucker decomposition is $R^n + n1R$. Moreover, CP decomposition usually achieves satisfactory results with a much larger rank; thus, it cannot better preserve the correlation of the nonlocal mode. Although LLRT is a simple and efficient method, it only employs the low-rank assumption in the nonlocal mode. Therefore, it provides motivation to search for another tensor representation method for balancing the correlations among all the modes to improve the performance of previous nonlocal-related methods. Recently, a new tensor-ring (TR) decomposition was proposed to represent a high-order tensor as a sequence of cyclically contracted third-order tensors [38], [39]. Compared with the Tucker decomposition, the number of
variables of TR decomposition is $\text{Inv}^2$ if the TR rank is $[r, \ldots, r]$ (in general, $r < R$). This is much smaller than that of Tucker decomposition. Moreover, because TR factors can be circularly shifted and treated equivalently, it is possible to balance effectively the correlations of all the modes. Based on the efficient representation of a high-order tensor, TR-based methods obtained better results in tensor completion [40]–[43] than those with the Tucker and CP decompositions.

Inspired by the effectiveness and superiority of TR representation, a novel model for HSI denoising through nonlocal TR (NLTR) decomposition was proposed in this article. By introducing the TR decomposition to represent the nonlocal grouped tensor, it was possible to capture and balance the global spectral correlation and nonlocal self-similarity correlation priors of the HSI more efficiently. The contributions of this article include the following.

1) A new tensor representation TR decomposition was introduced to depict the intrinsic spatial, spectral, and nonlocal correlation of similarly grouped tensors simultaneously. The advantage of TR, compared with the Tucker and CP decompositions, was determined by analysis. The new model could be easily extended to handle other high-dimension data-restoration problems.

2) An efficient optimization algorithm proximal alternating minimization (PAM) was employed as the TR factor solver. Extensive experimental results on both the simulated and real data sets demonstrate that the new NLTR method outperforms the state-of-the-art nonlocal methods for HSI denoising in terms of comparison of visual and quantitative indices. The framework of the proposed method is shown in Fig. 1.

The organization of this article is as follows. Section II introduces notations and an HSI denoising framework. The proposed NLTR decomposition and its optimization are formulated in Section III. Section IV reports the simulated and real experimental results and the parameter discussion. Finally, the conclusions for this article are presented in Section V.

II. NOTATIONS AND PROBLEM FORMULATION

A. Notations

In this article, lowercase or uppercase letters (e.g., $i, j \in \mathbb{R}$) are used to denote scalars. Vectors are represented by boldface lowercase letter, e.g., $\mathbf{x} \in \mathbb{R}^l$. Boldface capital letters (e.g., $\mathbf{X} \in \mathbb{R}^{I \times J}$) are employed to represent matrices. A tensor is a multi-dimensional data array, and an $n$-order tensor ($n \geq 3$) is represented by a calligraphic letter $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$. An element value of $\mathcal{X}$ in position $(i_1, i_2, \ldots, i_n)$ is represented as $\mathcal{X}(i_1, i_2, \ldots, i_n)$ or $x_{i_1i_2\ldots i_n}$. Two types of tensor matricization (also called unfolding) expressions are defined in this article. The normal mode-$k$ matricization of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ is introduced in [44] and represented as $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_{1:1:k-1}I_{k+1:1:n}}$. The other mode-$k$ matricization of $\mathcal{X}$, which is introduced in [38] for TR decomposition, is represented as $\mathbf{X}_{[k]} \in \mathbb{R}^{I_1I_{k+1:1:n}I_{k+1:1:n-1}}$. In contrast, the unfolding matrices $\mathbf{X}_{(k)}$ along the $k$-mode can be folded as a tensor by $\mathcal{X} = \text{fold}_k(\mathbf{X}_{(k)})$, and the same definition is used for the inverse operation of the second mode-$k$ matricization $\mathbf{X}_{(k)}$. Moreover, the Frobenius norm of a tensor $\mathcal{X}$ is calculated as $\|\mathcal{X}\|_F = \left(\sum_{i_1, i_2, \ldots, i_n} (x_{i_1i_2\ldots i_n})^2\right)^{1/2}$.

B. HSI Denoising via Tensor Decomposition

The denoising problem is a strong ill-posed problem. Mathematically, HSI denoising can be generally formulated as a regularization model

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{Y} - \mathcal{X}\|_F^2 + \lambda R(\mathcal{X}) \quad (1)$$

where $\mathcal{Y} \in \mathbb{R}^{M \times N \times B}$ is a noisy HSI with $M \times N$ spatial size and $B$ spectral bands, which is degraded by additive white Gaussian noise; $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$ is a clean HSI to be restored; and $\lambda$ is a positive regularization parameter used to balance these two terms. The first term is the data fidelity term, and $R(\mathcal{X})$ is the regularization term that characterizes the prior of $\mathcal{X}$.

With model (1), to better and efficiently restore $\mathcal{X}$ from $\mathcal{Y}$, the key problem is transformed to formulate an appropriate regularization term for $\mathcal{X}$. Because the HSI is a 3-D tensor compared with the 2-D image, many tensor-based methods perform HSI denoising by using the tensor decomposition strategy for the HSI data. This includes using Tucker decomposition [25] and CP decomposition [1] to approximate a low-rank tensor. Tucker decomposition is used to find a core tensor and a set of factor matrices, and to formulate them as

$$\mathcal{X} = S \times_1 U_1 \times_2 U_2 \times_3 U_3 \quad (2)$$

where $S$ is the core tensor and $U_k$ ($k = 1, 2, 3$) are the factor matrices. In addition, CP decomposition is used to decompose a tensor into the sum of rank-1 tensors and can be formulated as

$$\mathcal{X} = A \circ B \circ C = \sum_{k=1}^{R} A^{(k)} \circ B^{(k)} \circ C^{(k)} \quad (3)$$

where $A$, $B$, and $C$ are the factor matrices and $\circ$ is the vector out product.

These methods make use of the correlation of different bands in an HSI and can reduce its spatial–spectral redundancy. However, when the tensor representation was directly applied to the whole HSI data, it could not be used to explore the nonlocal self-similarity prior [32]. Recently, several researchers have confirmed that nonlocal methods can achieve the state-of-the-art results for HSI denoising [31]–[33], [35]. Similar to the whole HSI denoising regularization model, the nonlocal HSI noise removal model can be optimized by using the following model:

$$\min_{\mathcal{X}_i} \frac{1}{2} \|\mathcal{Y}_i - \mathcal{X}_i\|_F^2 + \lambda R(\mathcal{X}_i) \quad (4)$$

where $\mathcal{X}_i \in \mathbb{R}^{P^2 \times B \times K}$ is a third-order tensor stacked by nonlocal similar patches in the $i$-th exemplar patch. $P$ denotes the exemplar patch size and $K$ denotes the number of nonlocal similar patches. After all nonlocal groups $\mathcal{X}_i$ have recovered from $\mathcal{Y}_i$, then $\mathcal{X}_i$ can be aggregated to reconstruct a clean HSI $\mathcal{X}$. 

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
Similar to the model (1), the key problem is the exploring of the prior of $X$ and designing the related regularization to the optimized model (4). By grouping the similar patches together, the global correlation is explored not only in the spectral mode but also in the nonlocal mode. To explore these correlations, TDL [31], ITS [32], [35], and unidirectional low-rank tensor recovery [33] were proposed to describe the prior of $X$ and achieved the state-of-the-art results. TDL is based on low-rank tensor Tucker decomposition and ITS encodes sparsity insights delivered by low-rank tensor Tucker and CP decomposition.

### III. Proposed Nonlocal Tensor-Ring Decomposition for HSI Denoising

Although the mentioned nonlocal tensor HSI denoising methods achieved satisfactory results, there is still room for improvement of these results. Recently, a TR decomposition that represents a high-order tensor as a sequence of cyclically contracted third-order tensors was proposed by Zhao et al. [38]. Compared with Tucker and CP decompositions, TR decomposition can better approximate a high-order tensor, because each tensor factor can be circularly shifted and regarded equivalently under trace operation.

#### A. Tensor-Ring Decomposition

Compared with other forms of tensor decomposition, the TR form is a more general and effective decomposition. The purpose of TR decomposition is to represent a high-dimensional tensor by circular multilinear products on a series of third-order factor tensors (also called TR factors) [38]. Assume that $Z$ is an $n$-dimension tensor with the size of $I_1 \times I_2 \times \cdots \times I_n$. The representation of the TR decomposition of $Z$ is to search $n$ latent third-order core tensors $G = \{G(1), G(2), \ldots, G(n)\}$, where $G(k) \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$. In this case, each elementwise relation of $Z$ with the core tensor $G$ can be expressed by

$$Z(i_1, i_2, \ldots, i_n) = \text{Tr}(G^{(1)}(i_1)G^{(2)}(i_2) \cdots G^{(n)}(i_n))$$

where $G^{(k)}(i_k)$ is the $i_k$ lateral slice matrix of $G^{(k)}$ and $\text{Tr}(\cdot)$ denotes the matrix trace operation. According to the trace operation, the product of these slice matrices should be a square matrix; thus, TR decomposition sets the dimensions of the first and last cores to $r_1 = r_{n+1}$. In this case, vector $r = [r_1, r_2, \ldots, r_n]$ is called the TR-rank. The illustration of TR decomposition is shown in Fig. 2.

**Definition 1 (Tensor Multilinear Product [43]):** Let $G^{(k)} \in \mathbb{R}^{r_k \times I_1 \times I_2 \times \cdots \times I_{k-1}}$, $k = 1, \ldots, n$ are $n$ third-order tensors in the TR decomposition, and then, the multilinear product between $G^{(k)}$ and $G^{(k+1)}$ is denoted as $G^{(k+1)}$ and calculated as

$$G^{(k+1)}((j_k - 1)I_k + i_k) = G^{(k)}(i_k)G^{(k+1)}(j_k)$$

for $i_k = 1, \ldots, I_k$, $j_k = 1, \ldots, I_{k+1}$ and where $G^{(k+1)} \in \mathbb{R}^{r_k \times I_{k+1} \times I_{k+1}}$.

From Definition 1, it is possible to obtain the multilinear product of all the TR core tensors $G = \prod_{k=1}^{n} G^{(k)} = G^{(1)} \cdots G^{(n)}$ and the operator is defined by

$$Z(i_1, i_2, \ldots, i_n) = \text{Tr}(G(; i_1 + (i_2 - 1)I_1 \ldots + (i_n - 1)I_2 \cdots I_{n-1}, :))$$

**Lemma 1 (Circular Dimensional Permutation Invariance [38]):** For any $n$-order tensor $Z \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$, if $Z = \Phi(G^{(1)}, G^{(2)}, \ldots, G^{(n)})$ is its TR decomposition, $\sum_k Z_k \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ circularly shifts the dimensions of $Z$ by $k$. Then, the following relation can be obtained using:

$$\sum_k Z_k = \Phi(G^{(1)}, G^{(k+1)}, \ldots, G^{(n)}, G^{(1)}, \ldots, G^{(k-1)})$$

With Lemma 1, any TR core tensor can easily be shifted to the first position, indicating that TR decomposition has the ability to capture the correlation of an element in multiple dimensions.

#### B. Nonlocal Tensor-Ring Decomposition Model

From the definition and properties of TR decomposition, it is clear that it has numerous advantages over other tensor decomposition formats. First, the number of variables in storage is much smaller than that of Tucker decomposition. The number of variables in Tucker decomposition increases exponentially with respect to its original tensor order, indicating that the number of variables to be estimated for TR is obviously decreased. Thus, the estimates with TR will normally be more precise than for Tucker decomposition. Second, TR has the characteristics that each tensor factor can be circularly shifted and regarded equivalently under trace operation, whereas other tensor decomposition techniques fail to preserve this superiority. Third, the decomposition format of TR is tensor to tensor; thus, it has better ability to preserve the original data structure than Tucker or CP decomposition.

Furthermore, we focus on the 2-D spectral matrix analysis and illustrate the advantage of TR decomposition for HSI.
processing. We extract the spectral signatures of two different materials (denote \( S_1 \in \mathbb{R}^{B \times k} \) and \( S_2 \in \mathbb{R}^{B \times k} \)) from the HSI data set, where \( B \) is the number of band and \( k \) is the pixel number of the related materials. Then, we analyze the TR decomposition of the 2-D matrix \( S = [S_1, S_2] \in \mathbb{R}^{B \times 2k} \).

Typically, we adopt \( A_1 \in \mathbb{R}^{B \times r_1} \) and \( A_2 \in \mathbb{R}^{B \times r_1} \) to represent the basis of \( S_1 \) and \( S_2 \), respectively, i.e., \( S_1 = A_1 B_1 \) and \( S_2 = A_2 B_2 \). However, matrix/tensor decomposition can learn a global optimal basis of \( S \), ignoring the characteristics of each material. The TR decomposition of \( S \) is represented as \( S = \Phi(G^{(1)}, G^{(2)}) \), where \( G^{(1)} \in \mathbb{R}^{2k \times r_1} \) is the basis tensor. The core tensor \( G^{(1)} \) has the ability to learn the basis of each material separately. That is to say, TR ring decomposition can learn the basis of HSI’s different materials. Thus, TR decomposition has stronger ability than matrix/tensor decomposition to learn the basis of HSI’s different materials.

Based on the superiority of TR decomposition, HSI denoising can be achieved by finding the TR representation of the whole HSI. However, directly using TR decomposition for the whole HSI does not capture the nonlocal self-similarity; thus, TR decomposition was introduced for nonlocal HSI denoising. That is, the regularized term \( R(X_i) \) can be formulated in model (4) as \( \|X_i - \Phi([G_i])\|_F^2 \). By selecting an appropriate parameter, the proposed NLTR decomposition denoising model can be formulated as

\[
\min_{X_i,G_i} \frac{1}{2\gamma} \|Y_i - X_i\|_F^2 \quad \text{s.t.} \quad X_i = \Phi([G_i]).
\]

(10)

The proposed NLTR method mainly consists of three steps: grouping nonlocal similar fullband patches, low-rank TR approximation, and aggregating the estimated patches to a clean HSI. First, for each exemplar patch \( X_i \), all similar patches are searched in a local window by the Euclidean distance-based similarity metric. Second, the low-rank tensor approximation is performed by estimating the TR core tensors \( G_i \). Finally, the denoised HSI is recovered by aggregating all the patches. The three steps of the new method are presented in Fig. 1.

C. Optimization Method and Convergence Analysis

The optimization of the NLTR model can be divided into two steps. First, the TR cores \( G_i = \{G_i^{(1)}, G_i^{(2)}, G_i^{(3)}\} \) are estimated by solving the following optimization problem:

\[
\min_{G_i} \frac{1}{2}\|Y_i - \Phi([G_i])\|_F^2.
\]

(11)

Once the TR cores \( G_i \) are obtained, \( X_i \) is computed as \( X_i = \Phi([G_i]) \). The key problem is how to optimize the model (11). Clearly, the problem of (11) is not jointly convex for \( G_i \), but it is convex for \( G_i = \{G_i^{(1)}, G_i^{(2)}, G_i^{(3)}\} \) independently. To optimize the nonconvex problem effectively, PAM is applied [46]–[49].

Before optimizing (11), the problem of (11) is transformed into its equivalent form by introducing the proximal operator [50], and the problem can be rewritten as

\[
\min_{G_i} \frac{1}{2}\|Y_i - \Phi([G_i])\|_F^2 + \frac{\rho}{2}\|G_i - (G_i)^j\|_F^2
\]

(12)

where \( \rho \) denotes a positive proximal parameter and \( (G_i)^j \) is the result of the last iteration of \( G_i \). Then, it becomes possible to update alternately each core \( G_i^{(k)} \) for \( k = 1, 2, 3 \) as follows:

\[
\begin{align*}
(G_i^{(1)})^{j+1} &= \min_{G_i^{(1)}} \frac{1}{2}\|Y_i - \Phi([G_i^{(1)}, G_i^{(2)}, G_i^{(3)}])\|_F^2 \\
&\quad + \frac{\rho}{2}\|G_i^{(1)} - (G_i^{(1)})^j\|_F^2 \\
(G_i^{(2)})^{j+1} &= \min_{G_i^{(2)}} \frac{1}{2}\|Y_i - \Phi([G_i^{(1)}, G_i^{(2)}, G_i^{(3)}])\|_F^2 \\
&\quad + \frac{\rho}{2}\|G_i^{(2)} - (G_i^{(2)})^j\|_F^2 \\
(G_i^{(3)})^{j+1} &= \min_{G_i^{(3)}} \frac{1}{2}\|Y_i - \Phi([G_i^{(1)}, G_i^{(2)}, G_i^{(3)}])\|_F^2 \\
&\quad + \frac{\rho}{2}\|G_i^{(3)} - (G_i^{(3)})^j\|_F^2.
\end{align*}
\]

(13)

According to Lemma 1, each core tensor is shifted to the first position when optimizing this core tensor, which gives

\[
Z_{<k>} = (\vec{G}_k)_{<k>}(\vec{G}_{k+1}, \ldots, \vec{G}_{<k>})^T.
\]

(14)

Based on (14), the least square problem (13) can be solved as follows:

\[
\begin{align*}
(G_i^{(1)})^{j+1} &= \text{fold}_3 [(\vec{Y}_{<k>}(G_i^{(1)})^{<k>})^j + \rho (G_i^{(1)})^j] \\
(G_i^{(2)})^{j+1} &= \text{fold}_3 [(\vec{Y}_{<k>}(G_i^{(2)})^{<k>})^j + \rho (G_i^{(2)})^j] \\
(G_i^{(3)})^{j+1} &= \text{fold}_3 [(\vec{Y}_{<k>}(G_i^{(3)})^{<k>})^j + \rho (G_i^{(3)})^j]
\end{align*}
\]

(15)

where \( I \) is the identity matrix. After the optimization of core tensors \( G_i \), \( X_i \) can be updated as follows:

\[
X_i = \Phi([G_i]).
\]

(16)

The whole HSI denoising procedure is summarized in Algorithm 1.

According to the solver framework in [46], the proposed algorithm can converge to a critical point of the objection function. The theoretical proof of convergence is described in the Appendix.

D. Computational Complexity

Considering Algorithm 1 for an input-noisy HSI \( Y \in \mathbb{R}^{M \times N \times B} \), the total number of exemplar groups is \( S = O(MN) \). The main computation complexity of each exemplar group is to update \( G_i \) in Step 8. Assuming that the size of each group is \( \bar{Y}_i \in \mathbb{R}^{r_1 \times r_2 \times B} \) and the TR rank is set as \( r_1 = r_2 = r_3 = r \), the cost of each exemplar group updating in each iteration is \( O(r^6 + r^7 + r^{12}) \). For quite large spatial size, the computation cost is extremely high. However, the denoising can be performed on the \( S \) groups using parallel
Algorithm 1 NLTR Method for HSI Denoising

Require: Degraded HSI $\mathcal{Y}$, TR rank $r = [r_1, r_2, r_3]^T$, parameters $\rho$ and $\delta$.

1. Initialize: Let $\mathcal{X}^{(0)} = \mathcal{Y}$, $\mathcal{Y}^{(0)} = \mathcal{Y}$.
2. for $t = 1 : T$ do
3. Calculate $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \delta(\mathcal{Y} - \mathcal{X}^{(t-1)})$.
4. for each exemplar patch $Y_i$ do
5. Grouping non-local similar patches to form tensor $Y_i$.
6. Randomly initialize for core tensors $G_i$.
7. while not converged do
8. Alternately update $G_i = [G_i^{(1)}, G_i^{(2)}, G_i^{(3)}]$ via (15).
9. end while
10. Compute $X_i = \Phi([G_i])$.
11. end for
12. Aggregate all $X_i$ to achieve the denoised HSI $\mathcal{X}^{(T)}$.
13. end for

Ensure: Denoised HSI result $\mathcal{X}^{(T)}$.

In the experimental tests, simulated and real experiments were performed to illustrate the effectiveness of the NLTR method for HSI denoising. The following nine state-of-the-art and representative denoising methods were employed for comparison: bandwise BM3D [51], LRMR [52], BM4D [37], LRTA [25], PARAFAC [1], LRTDTV [28], TDL [31], KBR (an extension of ITS) [35], and LLRT [33]. The codes of the comparison methods are available from the authors’ homepages. Moreover, all parameter selections for the comparison approaches were determined from the authors’ codes or from the suggestion in their articles to obtain the best performance. To help reproduce all of the experimental results for HSI denoising, the NLTR code will be published on the author’s homepage.

IV. EXPERIMENTAL RESULTS

In the simulated experiments, three popular data sets were selected to simulate a noisy HSI. The first was the Columbia MSI data set, (CAVE5) which includes 32 real-world objects and materials. Each data set contained spatial size $512 \times 512$ and spectral dimension 31. In the test, five MSIs were selected from which to extract subimages with a size of $200 \times 200 \times 31$ to perform the experiments shown in Fig. 3. The second and third HSIs were selected from the hyperspectral Pavia City Center data set (PaC6) and the Washington DC Mall data set, (WDC7) respectively. Two subimages with size $200 \times 200 \times 50$ were extracted from these two hyperspectral data sets as the clean image. To simulate different noisy levels, zero-mean additive Gaussian noise was simulated with noise variance $\sigma$ of 10, 30, 50, or 100 for all the clean data sets.

To evaluate the denoising results of all methods, five objective quantitative indices were computed for comparison, including the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) [53], the feature similarity (FSIM) [54], the erreur relative globale adimensionnelle de synthese (ERGAS) [55], and the spectral angle mapper (SAM). PSNR and SSIM are two common indices for image restoration. FSIM illustrates the perceptual consistency compared with the ground truth. ERGAS and SAM are based on the spectral evaluation index. In general, larger PSNR, SSIM, and FSIM indicate better denoising results. In contrast, better results are needed to obtain smaller ERGAS and SAM. The image denoising results are shown for visual comparison.

1) Experimental Results for CAVE Data Set: Figs. 4 and 5 show the visual results for the CAVE Toy and Feather MSIs at two different noise intensities, respectively. LRMR failed to remove the noise in both the cases, because it used matrix low-rank to perform the denoising processing. With the noise variance $\sigma = 50$, other methods were able to remove the noise, but LRTA, PARAFAC, and LRTDTV produced

---

some artifacts. Moreover, BM3D and BM4D destroyed the image details. These visual conditions can be observed in the enlarged box. Fig. 5 shows the results of high noise variance. BM3D, LRMR, BM4D, LRTA, PARAFAC, and LRTDTV could not completely remove the noise completely, and the image textures were not preserved. The nonlocal-based methods TDL, KBP, and LLRT effectively eliminated noise, but image detail was lost. From the enlarged box, the new nonlocal-based TR decomposition method NLTR was found to achieve the best visual denoising results, thereby removing the noise effectively and preserving most of the image details and textures.

Table I presents the quantitative results from comparison of the denoising approaches with the CAVE data set at four different noise levels, calculated by averaging five MSI scenes for each noise level. We highlight the top three algorithms by using three different colors (red, green, and blue). Among the three colors, the best performance for each quality index is highlighted in red, the second-best performance is highlighted in green, and the third-best performance is highlighted in blue. From Table I, nonlocal-based tensor decomposition methods show better results than the other methods. Moreover, the NLTR approach achieves better improvement of the denoising results than the state-of-the-art nonlocal-based tensor methods TDL, KBP, and LLRT, which indicates the superiority of TR decomposition for tensor representation.

Fig. 6 shows the PSNR values of each band on the CAVE data set at four different noise levels, which are also computed by averaging the five MSIs. The results again demonstrate that nonlocal-based methods obtained superior results, and that the NLTR approach achieved the best results in most of the bands.

B. Experimental Results for the PaC and WDC Data Sets

Figs. 7 and 8 show the denoising results for the PaC and WDC data sets, respectively. BM3D removed the noise band by band, but the image details were destroyed.
TABLE I
QUANTITATIVE RESULTS OF DIFFERENT COMPARISON METHODS ON THE CAVE DATA SET UNDER DIFFERENT NOISE VARIANCES

<table>
<thead>
<tr>
<th>σ</th>
<th>Index</th>
<th>Noisy BM3D</th>
<th>LRM</th>
<th>BM4D</th>
<th>LRTA</th>
<th>PARAFAC</th>
<th>LRTDTV</th>
<th>TDL</th>
<th>KBR</th>
<th>LLRT</th>
<th>NLTR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>PSNR</td>
<td>28.132</td>
<td>39.415</td>
<td>38.725</td>
<td>42.789</td>
<td>39.917</td>
<td>32.653</td>
<td>39.498</td>
<td>42.597</td>
<td>43.803</td>
<td>45.074</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.539</td>
<td>0.963</td>
<td>0.928</td>
<td>0.981</td>
<td>0.952</td>
<td>0.846</td>
<td>0.969</td>
<td>0.979</td>
<td>0.984</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>FSIM</td>
<td>0.740</td>
<td>0.968</td>
<td>0.955</td>
<td>0.982</td>
<td>0.961</td>
<td>0.906</td>
<td>0.976</td>
<td>0.984</td>
<td>0.986</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>ERGAS</td>
<td>197.255</td>
<td>53.664</td>
<td>58.526</td>
<td>35.428</td>
<td>49.406</td>
<td>116.052</td>
<td>53.695</td>
<td>36.618</td>
<td>32.481</td>
<td>27.481</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.386</td>
<td>0.081</td>
<td>0.119</td>
<td>0.059</td>
<td>0.089</td>
<td>0.161</td>
<td>0.089</td>
<td>0.057</td>
<td>0.048</td>
<td>0.040</td>
</tr>
<tr>
<td>30</td>
<td>PSNR</td>
<td>18.589</td>
<td>33.379</td>
<td>30.643</td>
<td>36.459</td>
<td>33.785</td>
<td>32.163</td>
<td>34.163</td>
<td>36.399</td>
<td>38.359</td>
<td>38.854</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.171</td>
<td>0.880</td>
<td>0.687</td>
<td>0.932</td>
<td>0.857</td>
<td>0.795</td>
<td>0.905</td>
<td>0.935</td>
<td>0.955</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>FSIM</td>
<td>0.441</td>
<td>0.912</td>
<td>0.831</td>
<td>0.949</td>
<td>0.904</td>
<td>0.878</td>
<td>0.937</td>
<td>0.952</td>
<td>0.965</td>
<td>0.968</td>
</tr>
<tr>
<td></td>
<td>ERGAS</td>
<td>591.764</td>
<td>105.908</td>
<td>147.108</td>
<td>72.846</td>
<td>98.315</td>
<td>121.613</td>
<td>101.266</td>
<td>73.738</td>
<td>59.503</td>
<td>56.046</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.748</td>
<td>0.153</td>
<td>0.263</td>
<td>0.124</td>
<td>0.148</td>
<td>0.194</td>
<td>0.162</td>
<td>0.099</td>
<td>0.079</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.090</td>
<td>0.817</td>
<td>0.497</td>
<td>0.877</td>
<td>0.780</td>
<td>0.620</td>
<td>0.841</td>
<td>0.896</td>
<td>0.929</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>FSIM</td>
<td>0.322</td>
<td>0.873</td>
<td>0.723</td>
<td>0.920</td>
<td>0.865</td>
<td>0.786</td>
<td>0.901</td>
<td>0.926</td>
<td>0.947</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>ERGAS</td>
<td>986.273</td>
<td>139.467</td>
<td>231.920</td>
<td>102.019</td>
<td>134.111</td>
<td>174.271</td>
<td>131.471</td>
<td>130.699</td>
<td>80.879</td>
<td>81.344</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.942</td>
<td>0.192</td>
<td>0.380</td>
<td>0.176</td>
<td>0.181</td>
<td>0.301</td>
<td>0.211</td>
<td>0.121</td>
<td>0.104</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.023</td>
<td>0.708</td>
<td>0.262</td>
<td>0.747</td>
<td>0.643</td>
<td>0.329</td>
<td>0.724</td>
<td>0.801</td>
<td>0.841</td>
<td>0.874</td>
</tr>
<tr>
<td></td>
<td>FSIM</td>
<td>0.199</td>
<td>0.816</td>
<td>0.551</td>
<td>0.860</td>
<td>0.796</td>
<td>0.612</td>
<td>0.829</td>
<td>0.868</td>
<td>0.894</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>ERGAS</td>
<td>1972.346</td>
<td>201.051</td>
<td>433.415</td>
<td>157.884</td>
<td>204.139</td>
<td>317.231</td>
<td>191.158</td>
<td>150.094</td>
<td>136.462</td>
<td>118.671</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>1.184</td>
<td>0.269</td>
<td>0.608</td>
<td>0.270</td>
<td>0.244</td>
<td>0.493</td>
<td>0.297</td>
<td>0.168</td>
<td>0.175</td>
<td>0.142</td>
</tr>
</tbody>
</table>

The noise was eliminated by BM4D and KBR, but they obtained blurry results. Although LRTA, PARAFAC, LRTDTV, and TDL obtained acceptable results at low noise levels, they could not remove noise at high noise levels. In comparison, LLRT and the proposed NLTR method achieved superior results competitive with the other methods.

Tables II and III show the comparison of quantitative results for the PaC and WDC data sets, respectively. From Tables II and III, the NLTR method shows a slight improvement over the comparison methods at low noise levels. However, when the HSIs are seriously degraded by high noise levels, the proposed method demonstrates significant improvement in these four quantitative indices. Moreover, it can again be confirmed that the nonlocal-based methods outperformed the other methods, because they could explore more correlation information in the image. Fig. 9 shows the PSNR value of each band for the PaC and WDC data sets. It is evident that NLTR achieves much higher PSNR for almost every band except σ = 30 in the WDC data set, which indicates the robustness of the proposed method. In summary, extensive simulation experiments demonstrated that the NLTR method outperforms existing denoising approaches.

C. Real Experiments

In the above, a series of simulated experiments was performed to illustrate the effectiveness of the NLTR method; however, in real scenes, noise degradation is very complex. To demonstrate the effectiveness of the NLTR method in handling actual noisy HSIs, two remarkable noisy data sets were selected for testing. The first data set was collected by the NASA AVIRIS instrument over the Indian Pines test site,8 and the original size was 145 × 145 × 220. Some bands of this data set are seriously degraded with complicated noise. The second data set was collected by the reflective optics system imaging spectrometer (ROSIS) of the Pavia University,9 and the data contained 610 × 340 spatial pixels with 103 spectral bands. Because the first few bands are seriously degraded by noise, several bands with spatial size 340 × 340 were extracted to test.

Figs. 10 and 11 show the denoised results on the Indian Pines and Pavia University data sets, respectively. As presented

8https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.html
9http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes
in the denoised results of Fig. 10, it can be seen that the BM3D method blurred the image to remove the noise. With LRMR, LRTA, and TDL, the results included persistent noise in the images. PARAFAC and LRTDTV removed the noise but introduced some artifacts. Compared with the other methods, KBR and LLRT obtained better results, but some details were not well preserved. In contrast, NLTR simultaneously removed the noise and preserved the image details. The Pavia University data set was degraded by lower noise levels than with the Indian Pines data set, but BM3D, BM4D, LRTA, and TDL could not completely remove the noise. TDL achieved satisfactory denoising results in the simulation experiments, but failed to achieve such results using real data. LRMR effectively removed the noise using noise-free band information, but the image contrast was changed. Although LRTDTV and LLRT obtained better results than the above, the image details were blurred (see the enlarged box). After denoising, KBR and NLTR provided the best denoising results compared with all other methods. In summary, even without the ground-truth HSI as a reference, visual comparison also demonstrated the superiority of the proposed NLTR method for denoising real data. Moreover, we employed one non-reference image-assessment index Q-metric [56] to evaluate the results of real data sets. Table IV lists the nonreference index values of the real data sets. Combining the visual comparison and the nonreference image-assessment index, we can observe that the proposed method obtains the best results.
TABLE III
COMPARISON OF QUANTITATIVE RESULTS OF DIFFERENT METHODS USED ON THE WDC DATA SET UNDER DIFFERENT NOISE VARIANCES

<table>
<thead>
<tr>
<th>Noise variance</th>
<th>Method</th>
<th>PSNR</th>
<th>SSIM</th>
<th>FSIM</th>
<th>ERGAS</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>BM3D</td>
<td>28.136</td>
<td>0.936</td>
<td>0.928</td>
<td>0.975</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>LMRMR</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>BM4D</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>LRTA</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>PARAFAC</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>LRTDTV</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>TDL</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>KBR</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>LLRT</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>NLTR</td>
<td>31.180</td>
<td>0.962</td>
<td>0.986</td>
<td>0.982</td>
<td>0.975</td>
</tr>
</tbody>
</table>

---

D. Discussion

In the NLTR model, there are several parameters that need to be determined before performing Algorithm 1, including patch size $P$, similar patch numbers $K$, proximal parameter $\rho$, iterative regularization parameter $\delta$, and TR rank $r = [r_1, r_2, r_3]$. For fair comparison, the iterative regularization parameter was set to $\delta = 0.1$ as introduced in [32] and [35] for all the experiments. Although there are many other parameters in the model, all are very robust and could be fixed in all experiments. Next, the sensibility of these parameters will be analyzed. A subimage of Toy from the CAVE data set and the PaC data set was selected as the experimental object, and PSNR was employed as the evaluation measure to analyze the influence of the parameters.

1) Analysis of Patch Size $P$: Fig. 12(a) shows the PSNR values with different patch sizes and noise levels.
With different noise levels, the change in the PSNR values is not very rapid except for the cases when noise variances are equal to 30 and 50 in the PaC data set, which indicates the relative robustness of the parameter $P$. To balance the performance and efficiency, the patch size was set to $P = 6$ in all the experiments. Moreover, the overlapped pixel was set to 1.

2) Analysis of the Number of Similar Patches $K$: The relationship of PSNR with similar patch numbers $K$ is shown in Fig. 12(b). From the results, with the increase in the patch number, the PSNR values are significantly improved. When the patch number reaches 200, the PSNR results remain stable; thus, the patch number was empirically set to $K = 200$ in all the experiments.

3) Analysis of the Proximal Parameter $\rho$: In the NLTR solver, a proximal term was employed to guarantee algorithm stability. Fig. 12(c) shows the PSNR curves with respect to different values of $\rho$ at four noise levels. From Fig. 12(c), it can be observed that the PSNR achieves obvious improvement with the increase in $\rho$. However, when $\rho$ continued to increase, the performance worsened, and the best results were achieved with $\rho = 1$. Based on these results, the empirical setting $\rho = 1$ was used in all the experiments.

4) Analysis of TR Rank: The TR rank is an important parameter in the proposed model, and it controls the correlation of the nonlocal grouped tensor. To simplify the selection of the TR rank, it was assumed that the TR ranks of the second and third dimensions were equal, i.e., that TR rank was set as $r = [r_1, r_2, r_2]$. Fig. 13 shows the changes in the PSNR values with different TR ranks at four noise levels. From Fig. 13, it can be observed that the PSNR increases with increasing TR rank at a low noise level $\sigma = 10$. In contrast, $r_2$ should be smaller when the noise level is higher. The reason is that smaller TR rank can suppress noise when the noise
level is high. To balance the robustness of the TR rank and efficiency of the denoising results, the TR rank was empirically set to \( r = [12, 8, 8] \) at a low noise level (\( \sigma = 10 \)). For \( \sigma = \{30, 50, 100\} \), the TR rank was set to \( r = [9, 3, 3] \). In the real data, because the Indian Pines data set was seriously degraded by noise, the TR rank was set low \( r = [9, 3, 3] \). For the Pavia University data set, the TR rank was set to \( r = [12, 8, 8] \). In summary, the related parameters in the proposed model were relatively robust, because they could be fixed in all experiments and, thereby, obtained a better denoising result.

In our experimental results, we compared our NLTR method with many nonlocal modeling-based methods to show the superiority of TR decomposition for HSI denoising. In the next, to show the contribution of nonlocal modeling in NLTR, we compared the results of NLTR with the method that directly applies TR decomposition (without the nonlocal strategy) to the whole image. Table V lists the quantitative comparison results of NLTR with TR on the PaC data set. It can be observed that the nonlocal strategy significantly improves the denoised performance, especially for high noise variance.
In the following, a convergence analysis of the algorithm is given briefly. First, the convergence condition of the algorithm is reviewed.

**V. Conclusion**

In this article, a new NLTR decomposition was proposed for HSI denoising. The TR representation was used to explore simultaneously the nonlocal self-similarity and spectral correlation in the HSI. To optimize the TR factor efficiently, the PAM algorithm was designed to optimize the model. Extensive comparative experimental testing of both simulated and real data sets, using a group of state-of-the-art HSI denoising methods, demonstrated the superiority of NLTR decomposition for HSI denoising. The TR representation was used to explore parallel processing tasks.

Although the proposed model can achieve good denoising performance, there is still room for improvement. For example, automatic TR rank parameter estimation should be considered in the future, which is expected to improve the application and denoising results. In addition, the combination of a TR network with deep learning ideas [57], [58] offers a promising approach for HSI denoising.

**APPENDIX**

In the following, a convergence analysis of the algorithm is given briefly. First, the convergence condition of the algorithm is reviewed.

**Lemma 2** ([46]): Let function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \cup +\infty \) be a proper lower semicontinuous function, and let \((x^j)_{j \in \mathbb{N}} \in \mathbb{R}^n\) be a sequence such that:

- **C1** (Sufficient decrease condition): For each \( j \in \mathbb{N} \), \( f(x^{j+1}) + c_1 \|x^{j+1} - x^j\|_2^2 \leq f(x^j) \) holds for a constant \( c_1 \in (0, +\infty) \);

- **C2** (Relative error condition): For each \( j \in \mathbb{N} \), \( \exists g^{j+1} \in \partial f(x^{j+1}) \) such that \( \|g^{j+1}\|_2 \leq c_2 \|x^{j+1} - x^j\|_2^2 \) holds for a constant \( c_2 \in (0, +\infty) \);

- **C3** (Continuity condition): There exists a subsequence \( \{x^{j_k}\}_{k \in \mathbb{N}} \) and \( \bar{x} \in \mathbb{R}^n \) such that \( x^{j_k} \rightarrow \bar{x} \) and \( f(x^{j_k}) \rightarrow f(\bar{x}) \), as \( k \rightarrow +\infty \).

If the function \( f \) has the Kurdyka–Lojasiewicz (KL) property [46] at \( \bar{x} \) specified in C3, then the sequence \((x^j)_{j \in \mathbb{N}}\) converges to \( x^j \rightarrow \bar{x} \); \( \bar{x} \) is a critical point of function \( f \); and the sequence \((x^j)_{j \in \mathbb{N}}\) has a finite length, i.e., \( \sum_{j=0}^{\infty} \|x^{j+1} - x^j\| < +\infty \).

**Theorem 1**: Assume that the sequence \((G_i)^j\) generated by the proposed algorithm is bounded, then it converges to a critical point of the objective function.

**Proof**: Let \( f(G_i) = (1/2)\|y_i - \Phi(G_i)\|_F^2 \); from this, it is easy to know that \( f \) is a \( C^1 \) function with Lipschitz continuous gradient. Moreover, \( f \) is a polynomial of \( G_i \) because of the definition of the Frobenius norm. Based on the condition that the polynomial operator is a semialgebraic function, \( f \) is semialgebraic.

According to the literature [59], a semialgebraic real-valued function \( f \) is a KL function, thereby \( f \) satisfying the KL property at \( \forall x \in \text{dom}(f) \). In addition, the proximal operator is introduced in the solver, which is the framework of the alternating minimizing procedure [see 46, eqs. (61)–(63)] with \( B_1 = \rho I \), \( \rho > 0 \). Thus, the iterative sequence \( G_i^j \) obtained by the PAM algorithm satisfies the conditions C1, C2, and C3 in Lemma 2. That is, the bounded sequences generated by the proposed algorithm converge to a critical point of \( f \).

**ACKNOWLEDGMENT**

The authors would like to thank the Editors and the anonymous reviewers for their constructive comments that helped to improve the quality of this article.


Yong Chen received the B.S. degree from the School of Science, East China University of Technology, Nanchang, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China. From 2018 to 2019, he was a Research Intern with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. His research interests include remote sensing image processing and sparse optimization.

Wei He (S’14–M’17) received the B.S. degree from the School of Mathematics and Statistics and the Ph.D. degree in surveying, mapping and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, in 2012 and 2017, respectively. He is currently a Researcher with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. His research interests include image quality improvement, remote sensing image processing and low rank representation, and deep learning.

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.

Naoto Yokoya (S’10–M’13) received the M.Eng. and Ph.D. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2010 and 2013, respectively. From 2013 to 2017, he was an Assistant Professor with the University of Tokyo. From 2015 to 2017, he was an Alexander von Humboldt Fellow, with the German Aerospace Center (DLR), Weßling, Germany, and the Technical University of Munich (TUM), Munich, Germany. He is currently the Unit Leader of the RIKEN Center for Advanced Intelligence Project, Tokyo, where he has been leading the Geoinformatics Unit since 2018. He has been a Visiting Associate Professor with the Tokyo University of Agriculture and Technology, Tokyo, since 2019. His research interests include image processing, data fusion, and machine learning for understanding remote sensing images, with applications to disaster management.

Dr. Yokoya received the First Place from the 2017 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest organized by the Image Analysis and the Data Fusion Technical Committee (IADF TC). He has been the Chair of the IEEE GRSS IADF TC since 2019 and also the Secretary of the IEEE GRSS All Japan Joint Chapter since 2018. He has been an Associate Editor of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS) since 2018.

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.