Hyperspectral Image Denoising Using Factor Group Sparsity-Regularized Nonconvex Low-Rank Approximation

Yong Chen , Ting-Zhu Huang , Wei He , *Member, IEEE*, Xi-Le Zhao ,

Hongyan Zhang , *Senior Member, IEEE*, and Jinshan Zeng

Abstract—Hyperspectral image (HSI) mixed noise removal is a fundamental problem and an important preprocessing step in remote sensing fields. The low-rank approximation-based methods have been verified effective to encode the global spectral correlation for HSI denoising. However, due to the large scale and complexity of real HSI, previous low-rank HSI denoising techniques encounter several problems, including coarse rank approximation (such as nuclear norm), the high computational cost of singular value decomposition (SVD) (such as Schatten p-norm), and adaptive rank selection (such as low-rank factorization). In this article, two novel factor group sparsity-regularized nonconvex low-rank approximation (FGSLR) methods are introduced for HSI denoising, which can simultaneously overcome the mentioned issues of previous works. The FGSLR methods capture the spectral correlation via low-rank factorization, meanwhile utilizing factor group sparsity regularization to further enhance the low-rank property. It is SVD-free and robust to rank selection. Moreover, FGSLR is equivalent to Schatten p-norm approximation (Theorem 1), and thus FGSLR is tighter than the nuclear norm in terms of rank approximation. To preserve the spatial information of HSI in the denoising process, the total variation regularization is also incorporated into the proposed FGSLR models. Specifically, the proximal alternating minimization is designed to solve the proposed FGSLR models. Experimental results have demonstrated that the proposed FGSLR methods significantly outperform existing low-rank approximation-based HSI denoising methods.

Index Terms—Factor group sparsity regularization, hyperspectral image (HSI) denoising, nonconvex low-rank approximation, proximal alternating minimization (PAM).

I. INTRODUCTION

HYPERSPECTRAL image (HSI), a high dimension dataset of the same scene collected by a few hundreds of narrow wavelengths, contains more plenty of spectral information and is widely applied in remote sensing applications, such as environment monitoring and mineral exploration [1]. Due to the fundamental limitations of hyperpsectral imaging sensors, such as photon effects, stochastic error of photocounting, and calibration error [2], the process of HSI imaging is accompanied by different types of noise pollution, mainly including Gaussian, impulse, dead lines, stripes noises. The existence of hybrid noises extremely deteriorates the performance of further applications, e.g., super-resolution [3] and classification [4]. Since the high cost of hardware-based noise suppression solutions, it is reasonable to develop an effective HSI denoising approach as a preprocessing step to help the subsequent applications.

During the past decades, many different techniques have been developed for HSI denoising. A straightforward idea is to simply extend the 1-D signal or 2-D image processing methods for HSI pixel-by-pixel [5] or band-by-band [6] denoising. However, these methods cannot excavate the high spectral correlation across all bands, which is inadequate to obtain satisfactory results. To overcome the deficiencies, a variety of improved approaches have been developed by considering the correlation in the HSI dataset. For example, Othman and Qian [7] developed a wavelet shrinkage method in the derivative-domain for HSI denoising; Chen and Qian [8] proposed denoising method by combining the wavelet shrinkage with principal component analysis, and a traditional total variation (TV) regularization in the spatial-spectral dimension was proposed for HSI denoising [9]. Additionally, the framework of sparse representation was used to exploit the spatial-spectral correlation for HSI denoising [10]. In recent years, the subspace-based methods [11] were widely used to effectively describe the spectral correlation and achieved better results for HSI denoising, including fast hyperspectral denoising [12] and nonlocal meets global method [13].

In particular, the aforementioned methods are mainly employed to handle one or two types of noise, which are potentially restricted for real noisy HSI. There are various kinds of noise in HSI during the imaging process,

such as Gaussian, impulse, dead lines, and stripes noises. More recently, motivated by the effectiveness of deep convolutional neural networks (DCNNs) for strong spatial feature extraction and expression ability [14]–[16], DCNN-based methods were developed for HSI denoising [17]–[19]. Unfortunately, DCNN is hard to predict the spectral correlation. However, the huge spectral correlation has been proven effective to boost the denoising results [13]. As we know, the spectral correlation can be explored by rank minimization, thus low-rank approximation methods for HSI denoising are still the mainstream.

The robust principal component analysis (RPCA) technique was first introduced in [20] for HSI mixed noise removal. The underlying idea of RPCA is based on these two facts that the high spectral correlation of HSI can be captured by low-rank prior, meanwhile impulse, dead lines, and stripes noises possess the sparse characteristic. Under the framework of RPCA, Zhang et al. [20] first developed a low-rank matrix recovery (LRMR) approach for HSI mixed noise removal. However, the RPCA model cannot preserve the spatial information in HSI, leading to suboptimal denoising results for heavy noisy scenes. Thereupon, He et al. [21] incorporated band-by-band TV regularization into the RPCA framework to preserve the spatial piecewise smoothness and boost the denoising results. To improve the performance of band-by-band TV regularization, spatial-spectral TV (SSTV) and nonlocal self-similarity regularizations were widely introduced to the RPCA framework for HSI denoising [12], [22], [23]. The RPCA-based methods mainly devote to the selection of rank minimization, which is, in general, an NP-hard problem.

In order to break the computational hurdle brought by the rank minimization, most of the researchers devoted to replacing the rank minimization with a tractable surrogate. One of the most important alternatives is the nuclear norm, which is a convex surrogate of rank minimization and has been adopted in HSI denoising, such as LRMR [20], and TV regularized low-rank matrix factorization [21]. To improve the performance of nuclear norm, the weighted nuclear norm was proposed in [22] for HSI denoising. Another type of rank surrogate in HSI denoising is the nonconvex rank approximation function, which can adaptively shrink the singular values of low-rank objects. One popular nonconvex rank approximation function for HSI denoising is the Schatten p-norm with $(0 < p < 1)$ [24], [25]. Besides this, there are many other nonconvex rank approximation functions applied for HSI denoising, including nonconvex γ-norm approximation [26], and nonconvex normalized ϵ-penalty [27]. Although the above-mentioned rank surrogate approximations have been shown to be effective in the literature, yet all of them need to perform singular value decomposition (SVD) in each iteration, which causes the high cost of HSI denoising. In order to overcome this challenge, the idea of low-rank factorization was employed in the literature [28]–[33] to capture the high correlation of HSI. Benefit from the low-rank factorization, the computational efficiency is significantly improved. Due to the intrinsic tensor form of HSI, such low-rank matrix factorization methods were extensively extended to the tensor case

for HSI denoising, including t-SVD [34]–[36] Tucker decomposition [37]–[41], CANDECOMP/PARAFAC(CP) decomposition [42], and tensor ring decomposition [43]. However, the performance of these methods based on the idea of low-rank factorization in general heavily depends on the choice of rank, which is usually hard to determine in practice.

In summary, the existing low-rank approximations employed to preserve the globe spatial-spectral correlation are either not a tight rank approximation [20], [21], or computational cost [22], [24], or sensitive to rank selection [31], [38], [44]. In this article, we introduce a factor group sparsity-regularized nonconvex low-rank approximation (FGSLR) method for HSI denoising. The advantages of the proposed FGSLR methods can be summarized as follows. First, FGSLR approximates rank minimization on the basis of low-rank factorization, thus the optimization is of low computation cost. Second, the rank minimization of FGSLR is transformed to minimize the number of nonzero columns of factor, which can overcome the sensibility of the rank selection of previous low-rank factorization. Third, FGSLR is equivalent to the Schatten p-norm, thus it is a tighter rank approximation. Moreover, we incorporate the TV regularization into the factor to preserve the spatial information of HSI. The contributions of this article are summarized as follows.

1) A novel nonconvex low-rank approximation that uses factor group sparsity regularization is introduced for HSI denoising. This method could more effectively to exploit the low-rank characteristic of HSI and can integrate the superiorities and avoid the deficiencies of Schatten p-norm and low-rank factorization.

2) An efficient optimization framework proximal alternating minimization (PAM) is designed to minimize the nonconvex FGSLR models. Extensive experimental results with both the simulated and real-world datasets illustrate that the proposed FGSLR methods outperform the state-of-the-art low-rank methods for HSI denoising in terms of comparison of quantitative and visual performance.

The rest of this article is organized as follows. Section II briefly introduces some notations and related low-rank approximations for HSI denoising. The proposed FGSLR methods and the corresponding PAM optimization are presented in Section III. Section IV reports the performance of both simulated and real experiments and model analysis. Finally, we conclude this article in Section V.

II. NOTATIONS AND RELATED WORKS

A. Notations

The notations are mainly defined from [45] in this article. Scalars, vectors, and matrices are denoted as lowercase or uppercase letters (i, $I \in \mathbb{R}$), boldface lowercase letter ($\mathbf{b} \in \mathbb{R}^I$), and Boldface capital letters ($\mathbf{B} \in \mathbb{R}^{I_1 \times I_2}$), respectively. An n-D high-dimensional data tensor is represented as calligraphic letter $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$, in which $b_{i_1,i_2,\dots,i_n}$ is its element in location $(i_1, i_2, \dots, i_n)$. Since HSI is a 3-D tensor data, we mainly give the notation for 3-D tensor. The tube, row, and column fibers are vector defined as $\mathbf{x}_{i_1,i_2,:}$, $\mathbf{x}_{i_1,:,i_3}$ and $\mathbf{x}_{:,i_2,i_3}$

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

CHEN *et al.*: HSI DENOISING USING FGSLR

3

Notations	operations
$\mathbf{B}_{(n)}$ or $\mathcal{B}_{(n)}$	mode-n matricization of tensor $\mathcal{B}$: arranging the mode-n fibers as cloulumns.
$\text{Fold}_n(\mathbf{B}_{(n)})$	the inverse operator of mode-n matricization.
$<\mathcal{B}, \mathcal{A}>$	inner product of two tensors: $\sum_{i_1,i_2,\cdots,i_n} b_{i_1,i_2,\cdots,i_n} a_{i_1,i_2,\cdots,i_n}$.
$\|\mathcal{B}\|_F$	Frobenius-norm of tensor: $\sqrt{<\mathcal{B},\mathcal{B}>}$.
$\|\mathcal{B}\|_1$	ℓ_1-norm of tensor: $\sum_{i_1,i_2,\cdots,i_n} |b_{i_1,i_2,\cdots,i_n}|$.
$\|\mathcal{X}\|_{2,1}$	$\ell_{2,1}$-norm of 3-D tensor: $\sum_{i_3} \|\mathbf{X}_{:,:,i_3}\|_F$.
$\|\mathbf{B}\|_{2,1}$	$\ell_{2,1}$-norm of matrix: $\sum_{i_2} \|\mathbf{x}_{:,i_2}\|_2$.
$\mathcal{Y} = \mathcal{B} \times_k \mathbf{A}$	mode-k multiplication of $\mathcal{B}$: $\mathcal{Y}_{i_1,\cdots,i_{k-1},j,i_{k+1},\cdots,i_n} = \sum_{i_k} b_{i_1,i_2,\cdots,i_n} a_{j,i_k}$.
$\mathcal{Y} = \mathcal{B} \times_k \mathbf{A}$	matrix representation of $\mathcal{Y}$: $\mathbf{Y}_{(k)} = \mathbf{A}\mathbf{B}_{(k)}$.

in 3-D tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, respectively. Moreover, 3-D tensor $\mathcal{X}$ contains frontal, lateral, and horizontal slices denoted $\mathbf{X}_{:,:,i_3}$, $\mathbf{X}_{:,i_2,:}$, and $\mathbf{X}_{i_1,:,:}$, respectively. For the operations of tensor involved in this article, we summarize them in Table I.

B. Problem Formulation

Given a noisy HSI $\mathcal{Y} \in \mathbb{R}^{M \times N \times B}$, which is contaminated by various mixed noise, including Gaussian noise, impulse noise, stripes, and dead lines, where M and N are spatial sizes, and B is spectral number, HSI denoising is to restore the clean image $\mathcal{X}$ from the degraded $\mathcal{Y}$, and the degraded process can be regarded as an additive model

$$\mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N} \tag{1}$$

where $\mathcal{X}$, $\mathcal{S}$, and $\mathcal{N}$ are with the same size as $\mathcal{Y}$ and denote clean HSI, a mixture of impulse noise, stripes, and dead lines, and Gaussian noise, respectively.

Apparently, it is an ill-posed problem that directly estimating the clean HSI $\mathcal{X}$ from degraded model (1). To robustly restore the clean HSI, the maximum *a posteriori* (MAP) can be applied to obtain a regularization model used to address the ill-posed problem. Then, the HSI denoising model can be formulated as

$$\min_{\mathcal{X},\mathcal{S}} \frac{\beta}{2} \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + R_1(\mathcal{X}) + \tau R_2(\mathcal{S}) \tag{2}$$

where the first term is data fidelity term, and the last two terms are regularization terms specifying the prior information for variables $\mathcal{X}$ and $\mathcal{S}$ to be estimated. β, $\tau \geq 0$ are positive regularization parameters employed to balance their functions on the restored results. As the mixture of impulse noise, stripes and dead lines only corrupt a few parts or bands in HSI, they are regarded as sparse noise [20]. That is to say, the sparse regularization term [46], [47] of $R_2(\mathcal{S})$ can be consistently designed as $R_2(\mathcal{S}) = \|\mathcal{S}\|_1$, which is employed to isolate the sparse noise from the clean image. In the following, we should investigate the prior information and design effective regularization term for clean HSI $\mathcal{X}$.

C. Low-Rank Approximation for HSI Denoising

As we know, a linear combination of a small number of pure spectral endmembers can be used to approximately represent

each spectral signature for HSI, which indicates the high correlation among different spatial pixels across the spectral dimension [48]. Low-rank regularization is a strong tool for depicting the high correlation characteristic, thus the regularization term of $R_1(\mathcal{X})$ can be depicted as $R_1(\mathcal{X}) = \text{rank}(\mathbf{X}_{(3)})$, where $\text{rank}(\cdot)$ is the number of nonzero singular values. Due to the NP-hard problem of rank minimization, most works are motivated to develop different kinds of low-rank regularization [12], [20], [24], [38]. Below we review existed low-rank approximations that are employed to HSI denoising.

1) Nuclear Norm: Nuclear norm is a popular convex surrogate function for rank approximation, and it is widely applied to describe the high spectral correlation for HSI denoising [20], [21]. The nuclear norm is formulated as the sum of singular values of $\mathbf{X}_{(3)}$

$$\|\mathbf{X}_{(3)}\|_* = \sum_{i=1}^{\min(MN,B)} \sigma_i(\mathbf{X}_{(3)})$$

where $\sigma_i(\mathbf{X}_{(3)})$ is the ith singular value of $\mathbf{X}_{(3)}$. The main advantage of nuclear norm is that it is a convex function, but the rank cannot be well approximated in practice since all singular values are treated equally and concurrently minimized.

To improve the performance of nuclear norm and address the drawback, the weighted nuclear norm was proposed to approximate rank and applied to HSI denoising [22], [49], which is formulated as $\|\mathbf{X}_{(3)}\|_{\omega,*} = \sum_{i=1}^{\min(MN,B)} \omega_i \sigma_i(\mathbf{X}_{(3)})$, where the weight $\omega_i = 1/(\sigma_i(\mathbf{X}_{(3)}) + \epsilon)$, in which ϵ is a small number employed to avoid singularities. Different weights setting to the singular values can improve the flexibility of nuclear norm minimization for rank approximation.

2) Schatten p-Norm: The another general class of rank approximation is the Schatten p-norm ($0 \leq p \leq 1$) [50], which is formulated as

$$\|\mathbf{X}_{(3)}\|_{S_p} = \left(\sum_{i=1}^{\min(MN,B)} \sigma_i^p(\mathbf{X}_{(3)}) \right)^{1/p}.$$

In the boundary cases, the Schatten p-norm is degraded to the nuclear norm and rank function when $p = 1$ and $p = 0$, respectively. For $0 < p < 1$, Schatten p-norm is a nonconvex surrogate approximation for rank function. Therefore, Schatten p-norm is the a rank approximation between the rank minimization and nuclear norm with $0 < p < 1$. Based on the better approximation for rank minimization, Xie *et al.* [24] employed weighted Schatten p-norm to describe the low-rank prior on spectral dimension for HSI denoising.

3) Low-Rank Factorization: Both nuclear norm and Schatten p-norm need to perform SVD in each iteration, which results in plenty of computation cost. To apply the SVD-free rank surrogate, low-rank factorization strategy can be used to describe the rank minimization, due to the fact that any low-rank matrix can be factorized into the product of two low-dimension matrices, i.e., $\mathbf{X}_{(3)} = \mathbf{AB}$, where $\mathbf{A} \in \mathbb{R}^{B \times r}$, $\mathbf{B} \in \mathbb{R}^{r \times MN}$, and $r \ll \min\{B, MN\}$. Since the strong low-rank prior occurs in the third (spectral) dimension, then we can rewrite the matrix factorization as tensor form as follows:

$$\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$$

where $\mathcal{B} \in \mathbb{R}^{M \times N \times r}$ is obtained by reshaping each row of $\mathbf{B}$ as an image. From the HSI unmixing perspective, $\mathbf{A}$ is endmember matrix, and $\mathcal{B}$ denotes relative endmember abundances. This class of low-rank surrogate can achieve satisfactory HSI denoising by choosing appropriate rank r. Due to the high-efficiency and effectiveness the low-rank factorization has been widely studied for HSI denoising [13], [23], [33], [51]. On the basis of low-rank matrix factorization, various low-rank tensor decompositions, including low-rank Tucker decomposition [38], [39], low-rank CP decomposition [42], t-SVD [52], [53], and low-rank tensor ring decomposition [43], [54], are extensively used for HSI restoration.

Besides the above classes of rank approximations, there are other nonconvex rank approximation functions for HSI denoising, such as nonconvex γ-norm approximation [26], nonconvex normalized ϵ-penalty [27]. Although Schatten p-norm and low-rank factorization were commonly applied to HSI denoising and achieved satisfactory results, they are subject to certain deficiencies and cannot combine all advantages of these low-rank approximations. First, using a popular algorithm solver to optimize Schatten p-norm and nuclear norm-based methods need to perform SVD in each iteration, which brings high computational cost because the scale of real HSI data is too large. Second, performing low-rank factorization with SVD-free for rank approximation needs to preestimate the rank r, and the final results are greatly dependent on r. Intuitively, our objective is to find a low-rank approximation that can derive the advantages of Schatten p-norm and low-rank factorization and remedy their deficiencies for HSI denoising.

III. PROPOSED FGSLR METHODS

In this section, on the basis of low-rank factorization, we propose a new low-rank approximation model for HSI denoising, formulated as

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{\beta}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + (\|\mathbf{A}\|_{2,0} + \gamma \|\mathcal{B}\|_{2,0}) + \tau \|\mathcal{S}\|_1$$

where $\|\mathbf{A}\|_{2,0}$ and $\|\mathcal{B}\|_{2,0}$ are the number of nonzero columns of $\mathbf{A}$ and nonzero frontal slices of $\mathcal{B}$, respectively, γ is the parameter to tradeoff the contribution balance between $\|\mathbf{A}\|_{2,0}$ and $\|\mathcal{B}\|_{2,0}$. Since the $\|\cdot\|_{2,0}$ norm minimization is a NP-hard and unstable problem, we can relax the $\|\cdot\|_{2,0}$ by its convex $\|\cdot\|_{2,1}$ counterpart, that is,

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{\beta}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + (\|\mathbf{A}\|_{2,1} + \gamma \|\mathcal{B}\|_{2,1}) + \tau \|\mathcal{S}\|_1. \tag{3}$$

Compared to the previous SVD-based low-rank approximations model, we adopt $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$ to explore the spectral correlation prior of HSI, and further introduce a novel factor group sparsity regularization to enhance the robustness of rank selection of low-rank factorization. The new approximation in model (3) is named FGSLR. In the following section, we will analyze the advantage of our FGSLR in detail.

A. Advantage Analysis of FGSLR

For the SVD-free low-rank factorization $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$, the rank of $\mathbf{X}_{(3)}$ is strictly controlled by the number of columns of $\mathbf{A}$ or the spectral dimension of $\mathcal{B}$, thus r should be precisely selected. However, in the real HSI, the precise rank is hard to obtain. Therefore, it's necessary to design a method that can select the rank r adaptively under the framework of low-rank factorization. Fortunately, we have the following fact:

$$\text{rank}(\mathbf{X}_{(3)}) := \min_{\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}} \|\mathbf{A}\|_{2,0} = \min_{\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}} \|\mathcal{B}\|_{2,0}$$
$$= \min_{\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}} \frac{1}{2}(\|\mathbf{A}\|_{2,0} + \|\mathcal{B}\|_{2,0}). \tag{4}$$

Since the hard optimization of $\ell_{2,0}$-norm, the $\ell_{2,1}$-norm is chosen as the convex relaxation. That is to say, the regularization $\|\mathbf{A}\|_{2,1} + \gamma \|\mathcal{B}\|_{2,1}$ can further enhance the low-rank representation of original image $\mathcal{X}$. Moreover, the proposed FGSLR can get rid of the precise rank selection, and achieve a more robust strategy by the factor group sparsity regularization.

Very recently, Fan *et al.* [55] showed that the factor group sparsity regularization $\|\mathbf{A}\|_{2,1} + \gamma \|\mathcal{B}\|_{2,1}$ is equivalent to Schatten p-norm.

Theorem 1 (Factored Factorization of the Schatten p-Norm [55]): For any matrix $\mathbf{Z} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{Z}) = r \leq \min(m, n)$ and $\alpha > 0$, there exist two factor matrices $\mathbf{U} \in \mathbb{R}^{m \times d}$ and $\mathbf{V} \in \mathbb{R}^{d \times n}$ with $r \leq d \leq \min(m, n)$, then the following factored factorization of the Schatten p-norm is satisfied:

$$\|\mathbf{Z}\|_{S_{1/2}}^{1/2} = \sum_{i=1}^{r} \sigma_i^{1/2}(\mathbf{Z}) = \min_{\mathbf{Z} = \mathbf{UV}} \frac{1}{2}\left(\|\mathbf{U}\|_{2,1} + \|\mathbf{V}^{\mathbf{T}}\|_{2,1}\right). \tag{5}$$

If the SVD of $\mathbf{Z}$ denotes as $\mathbf{Z} = \mathbf{U_Z S_Z V_Z^T}$, then when $\mathbf{U} = \mathbf{U_Z S_Z^{1/2}}$ and $\mathbf{V} = \mathbf{S_Z^{1/2} V_Z^T}$, equality holds in (5).

Guided by Theorem 1, our factor group sparsity regularization $\|\mathbf{A}\|_{2,1} + \gamma \|\mathcal{B}\|_{2,1}$ is equivalent to $\|\mathbf{Z}\|_{S_{1/2}}^{1/2}$ if $\gamma = 1$. By extending (5), when the group sparsity of $\mathcal{B}$ is replaced by Frobenius-norm and $\gamma = (\alpha/2)$, then it corresponds to $\|\mathbf{Z}\|_{S_{2/3}}^{2/3}$

$$\|\mathbf{Z}\|_{S_{2/3}}^{2/3} = \sum_{i=1}^{r} \sigma_i^{2/3}(\mathbf{Z}) = \min_{\mathbf{Z} = \mathbf{UV}} \frac{2}{3\alpha^{1/3}}\left(\|\mathbf{U}\|_{2,1} + \frac{\alpha}{2}\|\mathbf{V}\|_F^2\right) \tag{6}$$

when (6) holds, $\mathbf{U} = \alpha^{1/3} \mathbf{U_Z S_Z^{2/3}}$ and $\mathbf{V} = \alpha^{-1/3} \mathbf{S_Z^{1/3} V_Z^T}$. Without loss of generality, both low-rank approximations of (5) and (6) are called FGSLR. From the rank approximation (4) and Theorem 1, the advanteges of our proposed FGSLR can be summarized as follows.

1) *Tighter Rank Approximation:* Compared with the nuclear norm, FGSLR is corresponding to Schatten p-norm ($0 < p < 1$), thus FGSLR can yield a tighter approximation to the rank minimization than the nuclear norm.

2) *Lower Computation Cost:* The nuclear norm, Schatten p-norm, and other nonconvex low-rank approximated minimization problems need to implement SVD calculation at every iteration in most of the optimization algorithms. On the contrary, optimizing two FGSLR are

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

CHEN *et al.*: HSI DENOISING USING FGSLR 5

SVD-free operators, and the factor regularizers can be solved by soft-threshold shrinkage operator or small linear equations. The computation cost of SVD is $O(mn\min\{m, n\})$, while FGSLR is $O(mnd)$.

3) *Robust to Rank Selection:* Compared to the direct low-rank factorization, we employ group sparse regularization to the factors, thus some of the columns of $\mathbf{U}$ and $\mathbf{V}^T$ are rapidly forced to zero in the iterative process. That is to say, this surrogate can automatically reduce the number of nonzero columns and decrease the rank r dynamically. In contrast, the surrogates of low-rank factorization need to manually choose the number of columns.

B. FGSLR With TV Regularization

The advantages mentioned above inspire us to employ the FGSLR as the rank surrogate to replace other classes of rank approximation methods (nuclear norm, Schatten p-norm, and low-rank factorization). Consequently, we can rewrite the model (3) and propose two HSI denoising models by incorporating the frameworks of FGSLR (5) and (6) into regularization model (2) as follows:

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{\beta}{2}\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + (\|\mathbf{A}\|_{2,1} + \|\mathcal{B}\|_{2,1}) + \tau\|\mathcal{S}\|_1$$
(7)

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{\beta}{2}\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + \left(\|\mathbf{A}\|_{2,1} + \frac{\alpha}{2}\|\mathcal{B}\|_F^2\right) + \tau\|\mathcal{S}\|_1$$
(8)

where $\mathbf{A} \in \mathbb{R}^{B \times d}$, $\mathcal{B} \in \mathbb{R}^{M \times N \times d}$ with $r \le d \le \min(m, n)$ (r is the rank of $\mathbf{X}_{(3)}$).

The proposed models (7) and (8) can well capture the low-rank characteristic of HSI, but fail to explore the prior of spatial information. Images of the real world are piecewise smoothness, that is, they contain similar pixels at spatial nearby locations [21]. If $\mathbf{A}$ is the endmember matrix, then each frontal slice of $\mathcal{B}$ is the related endmember abundance. From another side, each frontal slice of $\mathcal{B}$ is the eigen-image when $\mathbf{A}$ is the orthogonal basis matrix [12]. Therein, the spatial smooth structure of HSI can be explored by regularizing $\mathcal{B}$. With the increase of iteration, the number of nonzero columns of $\mathbf{A}$ dynamically reduces in the FGSLR model, then $\mathbf{A}$ can be regarded as the basis matrix of $\mathcal{X}$ in the spectral dimension. Therefore, $\mathcal{B}$ also has a similar spatial structure associated with original HSI $\mathcal{X}$. We can design the regularization term for $\mathcal{B}$ to preserve the spatial information of HSI. The popular prior for $\mathcal{B}$ is nonlocal self-similar [12], [13], [23], but it is very sensitive for sparse noise. Based on the spatial piecewise smoothness of HSI, the TV regularization [56] is introduced to constrain $\mathcal{B}$ for preserving spatial information. The idea of using smoothness constraint for factor matrices is also employed in the framework of tensor decomposition [57], [58], which achieves satisfactory results. In summary, incorporating TV regularization into model (7) and (8), we propose two

novel FGSLR models for HSI denoising as follows:

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{\beta}{2}\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + (\|\mathbf{A}\|_{2,1} + \|\mathcal{B}\|_{2,1})$$
$$+ \lambda\|\mathcal{B} \times \mathbf{D}\|_1 + \tau\|\mathcal{S}\|_1 \qquad (9)$$

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{\beta}{2}\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + \left(\|\mathbf{A}\|_{2,1} + \frac{\alpha}{2}\|\mathcal{B}\|_F^2\right)$$
$$+ \lambda\|\mathcal{B} \times \mathbf{D}\|_1 + \tau\|\mathcal{S}\|_1 \qquad (10)$$

where λ is a regularization parameter, and $\|\mathcal{B} \times \mathbf{D}\|_1 = \left\|\begin{pmatrix} \mathcal{B} \times_1 \mathbf{D}_1 \\ \mathcal{B} \times_2 \mathbf{D}_2 \end{pmatrix}\right\|_1$. $\mathbf{D}_1 \in \mathbb{R}^{M \times M}$ and $\mathbf{D}_2 \in \mathbb{R}^{N \times N}$ are first-order difference matrices. The proposed models (9) and (10) are named as FGSLR$_{1/2}$ and FGSLR$_{2/3}$, respectively.

In our FGSLR$_{1/2}$ and FGSLR$_{2/3}$ models, the low-rank surrogate of spectral dimension is effectively described by FGSLR, and the spatial structure is preserved by TV regularization. To simplify the notation, FGSLR$_{1/2}$ and FGSLR$_{2/3}$ are collectively referred to FGSLR in some descriptions. Relatively speaking, compared to the previous low-rank approximation methods, the proposed FGSLR remedies many deficiencies, including high computation cost of SVD, a convex surrogate of rank approximation, and a sensitive to rank selection. Moreover, we design the TV regularization to the dimension-reduced factor $\mathcal{B}$, which is more efficient than other methods that directly using TV regularization to HSI itself. It is worth noting that although previous work [39], [51] also employed group sparsity regularization for HSI denoising, which are very different from our work. They used group sparsity regularization to explore the shared sparsity pattern of the spatial difference images. In our work, we introduce a nonconvex low-rank approximation for HSI denoising by using factor group sparsity regularization.

C. PAM-Based Optimization Algorithm

In this section, we design an efficient algorithm to solve the proposed FGSLR model. It is easy to see that FGSLR$_{1/2}$ has one more nonsmooth term than FGSLR$_{2/3}$, but they are the same framework. Thus, we only show how to optimize one model FGSLR$_{1/2}$, and the optimization of another model FGSLR$_{2/3}$ can be derived by analogy.

Intuitively, model FGSLR$_{1/2}$ is not jointly convex for both variables $\mathbf{A}$, $\mathcal{B}$, $\mathcal{S}$, but it is convex for each variable $\mathbf{A}$, $\mathcal{B}$, $\mathcal{S}$ independently. To solve the model FGSLR$_{1/2}$ effectively, we introduce PAM algorithm [59], [60] to optimize it. Based on the framework of PAM, we first introduce the proximal operator for each variable, and then alternately optimizing one variable while fixing others. Therefore, solving model (9) can be transformed to alternately optimize the following three subproblems:

$$\begin{cases} \mathbf{A}^{k+1} = \arg\min_{\mathbf{A}} f(\mathbf{A}, \mathcal{B}^k, \mathcal{S}^k) + \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^k\|_F^2 \\ \mathcal{B}^{k+1} = \arg\min_{\mathcal{B}} f(\mathbf{A}^{k+1}, \mathcal{B}, \mathcal{S}^k) + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2 \\ \mathcal{S}^{k+1} = \arg\min_{\mathcal{S}} f(\mathbf{A}^{k+1}, \mathcal{B}^{k+1}, \mathcal{S}) + \frac{\rho}{2}\|\mathcal{S} - \mathcal{S}^k\|_F^2 \end{cases}$$

where $f(\mathbf{A}, \mathcal{B}, \mathcal{S})$, $(\rho/2)\|\cdot\|_F^2$, and ρ are the problem of (9), proximal term, and positive proximal parameter, respectively.

In the following, we solve the three subproblems according to the next steps.

1) Step 1: We solve the block of $\mathbf{A}$ subproblem as follows:

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}} \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B}^k \times_3 \mathbf{A} - \mathcal{S}^k\right\|_F^2 + \|\mathbf{A}\|_{2,1}$$
$$+ \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^k\|_F^2. \quad (11)$$

Since $\ell_{2,1}$-norm is a nonsmooth term, we cannot directly obtain the close-form solution of $\mathbf{A}$. To solve problem (11) efficiently, the alternating direction method of multipliers (ADMM) [61], [62] is applied. By introducing auxiliary valuable $\mathbf{R}$, we can transform problem (11) as constrained minimization model

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}} \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B}^k \times_3 \mathbf{A} - \mathcal{S}^k\right\|_F^2 + \|\mathbf{R}\|_{2,1}$$
$$+ \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^k\|_F^2, \quad \text{s.t.} \quad \mathbf{R} = \mathbf{A}. \quad (12)$$

Then, the augmented Lagrangian function of (12) is

$$L_\mu(\mathbf{A}, \mathbf{R}, \mathbf{W}_1) = \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B}^k \times_3 \mathbf{A} - \mathcal{S}^k\right\|_F^2$$
$$+ \|\mathbf{R}\|_{2,1} + \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^k\|_F^2 + \frac{\mu}{2}\|\mathbf{R} - \mathbf{A} + \mathbf{W}_1/\mu\|_F^2 \quad (13)$$

where $\mathbf{W}_1$ and μ are Lagrange multipliers and positive penalty parameter, respectively. Therefore, the ADMM framework indicates that the solution of (13) can be solved by following problem.

a) Optimizing $\mathbf{R}$: The subproblem of $\mathbf{R}$ is solved by

$$\mathbf{R}^{i+1} = \arg\min_{\mathbf{R}} \|\mathbf{R}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{R} - \mathbf{A}^{k+1,i} + \mathbf{W}_1^i/\mu\right\|_F^2.$$

Let $\mathbf{A}^{k+1,i} - \mathbf{W}_1^i/\mu = \mathbf{C}$, then the close-form solution of each column of $\mathbf{R}^{i+1}$ is

$$\mathbf{R}^{i+1}(:, n) = \begin{cases} \dfrac{\|\mathbf{C}(:, n)\|_2 - \frac{1}{\mu}}{\|\mathbf{C}(:, n)\|_2}\mathbf{C}(:, n), & \text{if } \|\mathbf{C}(:, n)\|_2 > \dfrac{1}{\mu} \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$
$$(14)$$

b) Optimizing $\mathbf{A}$: The subproblem of $\mathbf{A}$ is solved by

$$\mathbf{A}^{k+1,i+1} = \arg\min_{\mathbf{A}} \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B}^k \times_3 \mathbf{A} - \mathcal{S}^k\right\|_F^2$$
$$+ \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^k\|_F^2 + \frac{\mu}{2}\left\|\mathbf{R}^{i+1} - \mathbf{A} + \mathbf{W}_1^i/\mu\right\|_F^2$$

and its solution is computed by

$$\mathbf{A}^{k+1,i+1} = \left(\beta\left(\mathbf{Y}_{(3)} - \mathbf{S}_{(3)}^k\right)\mathbf{B}_{(3)}^{k,T} + \rho\mathbf{A}^k\right.$$
$$\left. + \mu\left(\mathbf{R}^{i+1} + \mathbf{W}_1^i/\mu\right)\right)\left(\beta\mathbf{B}_{(3)}^k\mathbf{B}_{(3)}^{k,T} + (\rho+\mu)\mathbf{I}\right)^{-1}.$$
$$(15)$$

c) Updating $\mathbf{W}_1$: The update of $\mathbf{W}_1$ is solved by

$$\mathbf{W}_1^{i+1} = \mathbf{W}_1^i + \mu\left(\mathbf{R}^{i+1} - \mathbf{A}^{k+1,i+1}\right). \quad (16)$$

We summarize the whole optimized process for $\mathbf{A}$-subproblem in Algorithm 1.

Algorithm 1 ADMM Algorithm for $\mathbf{A}$ Subproblem

Input: $\mathcal{Y}$, $\mathcal{S}^k$, $\mathcal{B}^k$, $\mathbf{A}^k$, β, ρ, μ, ε, and $i_{\max}$.
1: Initialize: Let $\mathbf{R} = \mathbf{W}_1 = \mathbf{0}$, and $i = 0$.
2: **while** not coverged **do**
3:     $i = i + 1$.
4:     Update $\mathbf{R}$ via (14).
5:     Update $\mathbf{A}$ via (15).
6:     Update $\mathbf{W}_1$ via (16).
7:     Check the convergence condition $\frac{\|\mathbf{A}^{i+1} - \mathbf{A}^i\|_F}{\|\mathbf{A}^i\|_F} \le \varepsilon$ and $i < i_{\max}$.
8: **end while**
Output: $\mathbf{A}^{k+1}$.

---

2) Step 2: The solution of $\mathcal{B}$ subproblem can be obtained by the following minimization problem:

$$\mathcal{B}^{k+1} = \arg\min_{\mathcal{B}} \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{k+1} - \mathcal{S}^k\right\|_F^2$$
$$+ \|\mathcal{B}\|_{2,1} + \lambda\|\mathcal{B} \times \mathbf{D}\|_1 + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2. \quad (17)$$

Similar to $\mathbf{A}$ subproblem, we also apply ADMM to solve problem of (17) because two nonsmooth terms of $\ell_{2,1}$ and ℓ_1-norms. Two auxiliary valuables $\mathcal{C}$ and $\mathcal{G}$ are introduced to replace two nonsmooth terms, the problem of (17) is equivalent to the following problem:

$$\mathcal{B}^{k+1} = \arg\min_{\mathcal{B}} \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{k+1} - \mathcal{S}^k\right\|_F^2 + \|\mathcal{C}\|_{2,1}$$
$$+ \lambda\|\mathcal{G}\|_1 + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$\text{s.t.} \quad \mathcal{C} = \mathcal{B}, \quad \mathcal{G} = \mathcal{B} \times \mathbf{D}$$

whose augmented Lagrangian function is

$$L_\mu(\mathcal{B}, \mathcal{C}, \mathcal{G}, \mathcal{W}_2, \mathcal{W}_3)$$
$$= \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{k+1} - \mathcal{S}^k\right\|_F^2$$
$$+ \|\mathcal{C}\|_{2,1} + \lambda\|\mathcal{G}\|_1 + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$+ \frac{\mu}{2}\|\mathcal{C} - \mathcal{B} + \mathcal{W}_2/\mu\|_F^2 + \frac{\mu}{2}\|\mathcal{G} - \mathcal{B} \times \mathbf{D} + \mathcal{W}_3/\mu\|_F^2$$
$$(18)$$

where $\mathcal{W}_2$ and $\mathcal{W}_3$ are Lagrange multipliers, and based on the ADMM framework, we can solve the following subproblems via an iterative way.

a) Optimizing $\mathcal{C}$: The subproblem of $\mathcal{C}$ can be solved by group sparse soft-threshold shrinkage operator, then the frontal slices of $\mathcal{C}$ is updated by

$$\mathcal{C}^{j+1}(:, :, n)$$
$$= \begin{cases} \dfrac{\|\mathcal{H}(:, :, n)\|_F - \frac{1}{\mu}}{\|\mathcal{H}(:, :, n)\|_F}\mathcal{H}(:, :, n), & \text{if } \|\mathcal{H}(:, :, n)\|_F > \dfrac{1}{\mu} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$
$$(19)$$

where $\mathcal{H} = \mathcal{B}^{k+1,j} - \mathcal{W}_2^j/\mu$.

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

CHEN *et al.*: HSI DENOISING USING FGSLR
7

b) Optimizing $\mathcal{G}$: We can employ ℓ_1-norm soft-threshold shrinkage operator to obtain the solution as follows:

$$\mathcal{G}^{j+1} = \text{sign}(\mathcal{J}) \max\left(|\mathcal{J}| - \frac{\lambda}{\mu}, 0\right) \quad (20)$$

where $\mathcal{J} = \mathcal{B}^{k+1,j} \times \mathbf{D} - \mathcal{W}_3^j/\mu$.

c) Optimizing $\mathcal{B}$: The problem of $\mathcal{B}$ is a least squares minimization, and the solution can be obtained by

$$\beta \mathcal{B}^{k+1,j+1} \times_3 \left(\mathbf{A}^{k+1,T} \mathbf{A}^{k+1}\right) + (\rho + \mu) \mathcal{B}^{k+1,j+1}$$
$$+ \mu \mathcal{B}^{k+1,j+1} \times (\mathbf{D}^T \mathbf{D}) = \beta(\mathcal{Y} - \mathcal{S}^k) \times_3 \mathbf{A}^{k+1,T}$$
$$+ \rho \mathcal{B}^k + \mu \left(\mathcal{C}^{j+1} + \mathcal{W}_2^j/\mu\right) + \mu \left(\mathcal{G}^{j+1} + \mathcal{W}_3^j/\mu\right) \times \mathbf{D}^T$$
$$(21)$$

which is a Sylvester equation and can be efficiently solved by SVD (small matrix dimension: $d \times d$) and fast Fourier transformation operator (refer to [63] Theorem 1).

d) Updating $\mathcal{W}_2$ and $\mathcal{W}_3$: The update of these two Lagrange multipliers is calculated by

$$\begin{cases} \mathcal{W}_2^{j+1} = \mathcal{W}_2^j + \mu \left(\mathcal{C}^{j+1} - \mathcal{B}^{k+1,j+1}\right) \\ \mathcal{W}_3^{j+1} = \mathcal{W}_3^j + \mu \left(\mathcal{G}^{j+1} - \mathcal{B}^{k+1,j+1} \times \mathbf{D}\right). \end{cases} \quad (22)$$

In summary, the ADMM algorithm for $\mathcal{B}$ subproblem is presented in Algorithm 2.

Algorithm 2 ADMM Algorithm for $\mathcal{B}$ Subproblem

Input: $\mathcal{Y}$, $\mathcal{S}^k$, $\mathcal{B}^k$, $\mathbf{A}^{k+1}$, β, ρ, μ, λ, ε, and $j_{\max}$.
 1: Initialize: Let $\mathcal{C} = \mathcal{G} = \mathcal{W}_2 = \mathcal{W}_3 = \mathbf{0}$, and $j = 0$.
 2: **while** not coverged **do**
 3:    $j = j + 1$.
 4:    Update $\mathcal{C}$ via (19).
 5:    Update $\mathcal{G}$ via (20).
 6:    Update $\mathcal{B}$ via (21).
 7:    Update $\mathcal{W}_2$ and $\mathcal{W}_3$ via (22).
 8:    Check the convergence condition $\frac{\|\mathcal{B}^{j+1} - \mathcal{B}^j\|_F}{\|\mathcal{B}^j\|_F} \leq \varepsilon$ and $j < j_{\max}$.
 9: **end while**
Output: $\mathcal{B}^{k+1}$.

---

3) Step 3: Finally, the update of sparse noise $\mathcal{S}$ is solved by

$$\mathcal{S}^{k+1} = \arg\min_{\mathcal{S}} \frac{\beta}{2} \left\| \mathcal{Y} - \mathcal{B}^{k+1} \times_3 \mathbf{A}^{k+1} - \mathcal{S} \right\|_F^2$$
$$+ \tau \|\mathcal{S}\|_1 + \frac{\rho}{2} \|\mathcal{S} - \mathcal{S}^k\|_F^2 \quad (23)$$

which also can be solved by ℓ_1-norm soft-threshold shrinkage operator

$$\mathcal{S}^{k+1} = \text{sign}(\mathcal{L}) \max\left(|\mathcal{L}| - \frac{\tau}{\beta + \rho}, 0\right) \quad (24)$$

where $\mathcal{L} = (\beta(\mathcal{Y} - \mathcal{B}^{k+1} \times_3 \mathbf{A}^{k+1}) + \rho \mathcal{S}^k)/(\beta + \rho)$.

Summarizing the aforementioned solution of three block subproblems, we arrive at the PAM algorithm to solve the $\text{FGSLR}_{1/2}$ model (9) for HSI denoising, as presented in Algorithm 3. Moreover, the proposed Algorithm 3 can converge to a critical point of the objection function, which is demonstrated by the solver framework in [60].

Algorithm 3 PAM Algorithm for $\text{FGSLR}_{1/2}$ Model

Input: Noisy HSI $\mathcal{Y}$, rank $r \leq d \leq \min\{MN, B\}$, and $k_{\max}$.
 1: Initialize: $\mathbf{A} = \mathbf{U}\mathbf{S}^{1/2}$, $\mathcal{B} = \text{Fold}_3(\mathbf{S}^{1/2}\mathbf{V}^T)$ and $k = 0$, where $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ are the SVD of $\mathbf{Y}_{(3)}$.
 2: **while** not coverged **do**
 3:    $k = k + 1$.
 4:    Update $\mathbf{A}$ via Algorithm 1.
 5:    Update $\mathcal{B}$ via Algorithm 2.
 6:    Update $\mathcal{S}$ via (24).
 7:    Update $d = \text{nzc}(\mathbf{A})$, where $\text{nzc}(\mathbf{A})$ is the number of nonzero columns of $\mathbf{A}$.
 8:    Remove the zero columns and zero frontal slices of $\mathbf{A}$ and $\mathcal{B}$, respectively.
 9:    Check the convergence condition $\frac{\|\mathcal{B}^{k+1} \times_3 \mathbf{A}^{k+1} - \mathcal{B}^k \times_3 \mathbf{A}^k\|_F}{\|\mathcal{B}^k \times_3 \mathbf{A}^k\|_F} \leq \varepsilon$ and $k < k_{\max}$.
10: **end while**
Output: Denoised HSI $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$.

D. Analysis of Computational Complexity

To illustrate the high efficiency of the proposed method, we give the analysis of computational complexity on the PAM-based Algorithm 3. Suppose the size of input noisy HSI is $\mathcal{Y} \in \mathbb{R}^{M \times N \times B}$, and the initial rank is $r \leq d \leq \min\{MN, B\}$. The computational complexity is consists of three steps of $\mathbf{A}$, $\mathcal{B}$, and $\mathcal{S}$ update. For the subproblem of $\mathbf{A}$ on Algorithm 1, solving (15) of $\mathbf{A}$ subproblem is the main computation cost, which involves matrix multiplication with computation complexity $O(dMNB)$. Similarly, in step 2, the main computation cost of Algorithm 2 is to solve $\mathcal{B}$. Although it needs to apply the SVD operator, the dimension of the matrix is too small, thus it cannot consume much cost. The computation complexity of Sylvester equation of $\mathcal{B}$ is $O(d^2MN + dMN \log MN)$. Finally, the solution of $\mathcal{S}$ is a soft-threshold shrinkage operator, which has computation complexity $O(MNB)$. In summary, the total computation complexity of Algorithm 3 is $O(dMNB + d^2MN + dMN \log MN + MNB)$ at every iteration. It is worth noting that d is decreasing with the iteration increase. While the computation complexity of only one SVD operator for input HSI $\mathcal{Y}$ is $O(MNB \min\{MN, B\})$, thus the proposed method is lower computation cost than other low-rank approximation-based HSI denoising methods.

IV. Experimental Results and Discussions

In this section, extensive simulated and real experimental results are presented to demonstrate the performance of the proposed FGSLR methods. Seven representative low-rank approximation methods are considered for quantitative and visual comparison, i.e., LRTV [21] (nuclear norm and TV-based method), LSSTV [22] (weighted nuclear norm and SSTV-based method), WSNM [24] (Schatten p-norm-based method), NMoG [31] (low-rank factorization-based method), LRTDTV [38] (Tucker decomposition and SSTV-based method), TLR-SSTV [35] (t-SVD and SSTV-based method), and LRTDGS [39] (Tucker decomposition and group sparsity-based method). We conduct all experiments on MATLAB R2017b using a desktop of 16-GB RAM, with an

Intel Core i7-8700K CPU at 3.70 GHz. For all the competitors, the settings of parameters are adjusted as suggested in the literature, while for the proposed FGSLR, we will provide some careful discussions on the selection of parameters in the discussion part. The code of our method will be available at https://chenyong1993.github.io/yongchen.github.io/.

A. Simulated Experiments

To evaluate the noise removal effect of our FGSLR methods, we conduct simulated experiments by comparing both visual and quantitative performances. Two representative clean HSI datasets are employed to simulate noisy HSI, i.e., HYDICE Washington DC Mall data (WDC[1]) and hyperspectral Pavia City Center data (PaC[2]). Due to the big size of original data, we extract a subregion with the size of $256 \times 256 \times 191$ and $200 \times 200 \times 80$ for WDC, and PaC, respectively. There are several different types of noise in real HSI, including Gaussian, impulse, dead lines, and stripes noises. To simulate these scenarios as best as possible, we conduct five different noise cases based on the mixed noise setting of compared methods to thoroughly compare. Before the noise simulation, we normalize the gray value of each band into [0, 1].

Case 1 (Gaussian Noise): Zero-mean Gaussian noise with noise variance 0.15 is added to all bands.

Case 2 (Gaussian Noise + Impulse Noise): Mixtures of zero-mean Gaussian noise and impulse noise are added to all bands. The variances of Gaussian noise and percentages of impulse noise are randomly sampled within the range of [0.05, 0.2] and [0, 0.2], respectively.

Case 3 (Gaussian Noise + Impulse Noise + Dead Lines): Mixtures of zero-mean Gaussian noise, impulse noise, and dead lines are added to HSI. The variances of Gaussian noise are randomly sampled within the range of [0.05, 0.2], and the percentages of impulse noise are set as 0.05 to all bands. Moreover, 40% of bands are added dead lines whose numbers are from 5 to 12.

Case 4 (Gaussian Noise + Impulse Noise + Stripes): Mixtures of zero-mean Gaussian noise, impulse noise, and stripes are added to HSI. The distributions of Gaussian and impulse noises are the same as Case 3. Moreover, 40% of bands are added stripes whose number is from 5 to 12.

Case 5 (Gaussian Noise + Impulse Noise + Dead Lines + Stripes): Mixtures of zero-mean Gaussian noise, impulse noise, dead lines, and stripes are added to HSI. The distributions of Gaussian and impulse noises are the same as Case 2. Moreover, we, respectively, select 20% of bands in which to add dead lines and stripes whose numbers are all from 5 to 12. Furthermore, the number of overlapped bands of dead lines and stripes is set as 5% of bands.

Since the ground-truth HSI is given for the simulated experiments, we employ five quantitative picture indices for comparison, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM), feature similarity (FSIM), erreur relative global adimensionnelle de synthese (ERGAS), and spectral

[1] https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html
[2] http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

angle mapper (SAM) [39]. The five indices contain the evaluation of spatial and spectral information preservation, and the values of PSNR, SSIM, and FSIM are obtained by averaging all bands. The larger PSNR, SSIM, and FSIM are, and the smaller ERGAS and SAM are, the better HSI denoising result is.

1) Experimental Results on WDC Dataset: Table II lists the five quantitative indices comparison of different HSI denoising methods on the simulated WDC dataset. The best results among all methods are highlighted in bold in each case, and we underline the value as the second-best. From the table, we can see that the proposed FGSLR methods can significantly outperform other methods with respect to the five evaluation indices. Especially, $\text{FGSLR}_{1/2}$ achieves the best results in all cases, and the second-best results are obtained by $\text{FGSLR}_{2/3}$. Based on the better results of FGSLR than other methods, it demonstrates that the FGSLR is better to approach the rank minimization than other approximations. Although LRTDTV and LRTDGS employ the low-rank tensor decomposition to capture the high correlation of HSI, they cannot dynamically adjust the rank with the iteration increase, and the results are largely determined by the rank selection. Moreover, it is observed that $\text{FGSLR}_{1/2}$ achieves better results than $\text{FGSLR}_{2/3}$ in all cases, which also verifies the better rank approximation with the selection of a smaller p. In summary, our FGSLR methods can obtain the best quantitative results.

Due to the page limitation, we present the most complex noise Case 5 for visual comparison. Fig. 1 presents the denoising results of the simulated WDC dataset under Case 5, and the false color image is composed of bands 100, 9, and 7. As shown in Fig. 1(b), the image is seriously contaminated by a mixture of Gaussian, impulse, dead lines, and stripes noises. After deionised by all methods, the mixed noises are largely damped. However, LRTV, LSSTV, WSNM, and NMoG cannot thoroughly remove the noises, and some Gaussian noises left in the results as shown in the enlarged box of Fig. 1(c)–(f). Although LRTDTV successfully removes the mixture noise, it obtains the over-smoothed denoising results. TLR-SSTV cannot remove the mixed noise, especially for stripe and dead line noises. LRTDGS and $\text{FGSLR}_{2/3}$ perform satisfactory results, but image details are slightly destroyed shown in the enlarged box. From Fig. 1(j), we can observe that $\text{FGSLR}_{1/2}$ effectively removes the mixture noise as well as restore the image details and structures of the original image.

Fig. 2 shows the values of PSNR of each band provided by different methods on Cases 1–5. It can be seen that the proposed FGSLR methods achieve the higher PSNR values than other methods in almost all bands for every case, which verifies the robustness of the proposed methods on restoring all bands and also demonstrates the superiority of the nonconvex low-rank approximation compared with others.

2) Experimental Results on PaC Dataset: Five quantitative indices comparison on simulated PaC dataset under different noise cases are listed in Table III. From the results, we again find that our FGSLR methods achieve the best results in all compared methods on Cases 1–5, and $\text{FGSLR}_{1/2}$ also obtains the better results than $\text{FGSLR}_{2/3}$. Tensor low-rank-based methods LRTDTV and LRTDGS perform better results than

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

CHEN *et al.*: HSI DENOISING USING FGSLR

9

TABLE II

FIVE QUANTITATIVE INDICES COMPARISON OF DIFFERENT HSI DENOISING METHODS ON SIMULATED WDC DATASET

| Case	Index	Noisy	LRTV	LSSTV	WSNM	NMoG	LRTDTV	TLR-SSTV	LRTDGS	FGSLR$_{2/3}$	FGSLR$_{1/2}$
Case 1	PSNR	16.48	31.55	32.67	33.21	33.35	33.47	29.64	34.03	33.86	**34.38**
	SSIM	0.273	0.883	0.912	0.921	0.926	0.926	0.878	0.935	0.935	**0.942**
	FSIM	0.624	0.941	0.953	0.956	0.958	0.964	0.924	0.963	0.962	**0.966**
	ERGAS	582.687	103.586	86.751	81.956	80.749	79.281	122.740	74.269	75.441	**71.067**
	SAM	0.626	0.121	0.101	0.091	0.093	0.080	0.113	0.081	0.081	**0.076**
Case 2	PSNR	13.84	32.36	33.88	33.37	30.45	34.08	30.51	34.73	35.05	**35.64**
	SSIM	0.205	0.907	0.932	0.935	0.882	0.937	0.899	0.945	0.950	**0.955**
	FSIM	0.592	0.955	0.963	0.960	0.940	0.964	0.936	0.969	0.971	**0.974**
	ERGAS	842.601	94.790	76.220	158.970	121.139	74.338	112.867	69.005	66.114	**61.857**
	SAM	0.719	0.112	0.090	0.152	0.148	0.075	0.106	0.076	0.073	**0.068**
Case 3	PSNR	15.19	32.43	33.67	34.06	30.03	34.06	29.90	34.37	34.93	**35.54**
	SSIM	0.254	0.911	0.937	0.943	0.878	0.939	0.887	0.946	0.953	**0.959**
	FSIM	0.630	0.956	0.964	0.966	0.934	0.965	0.929	0.969	0.972	**0.975**
	ERGAS	689.707	95.535	78.545	84.850	137.989	75.384	121.936	75.545	67.311	**62.703**
	SAM	0.654	0.113	0.091	0.102	0.157	0.073	0.122	0.085	0.072	**0.067**
Case 4	PSNR	14.97	32.24	33.55	34.03	32.03	34.03	30.15	34.75	34.82	**35.43**
	SSIM	0.244	0.907	0.930	0.937	0.921	0.937	0.890	0.947	0.948	**0.955**
	FSIM	0.622	0.954	0.962	0.965	0.958	0.964	0.932	0.970	0.970	**0.973**
	ERGAS	702.842	96.266	78.605	84.429	115.671	74.668	116.837	68.421	67.746	**63.147**
	SAM	0.659	0.113	0.093	0.106	0.147	0.075	0.113	0.074	0.074	**0.068**
Case 5	PSNR	13.75	32.05	33.52	33.07	30.21	33.90	29.97	34.12	34.85	**35.03**
	SSIM	0.204	0.902	0.931	0.933	0.876	0.936	0.886	0.941	0.949	**0.954**
	FSIM	0.587	0.952	0.962	0.959	0.938	0.963	0.929	0.967	0.970	**0.972**
	ERGAS	861.064	99.323	79.581	170.515	123.200	76.140	120.241	80.396	67.510	**66.624**
	SAM	0.728	0.117	0.094	0.152	0.150	0.075	0.119	0.095	0.074	**0.071**
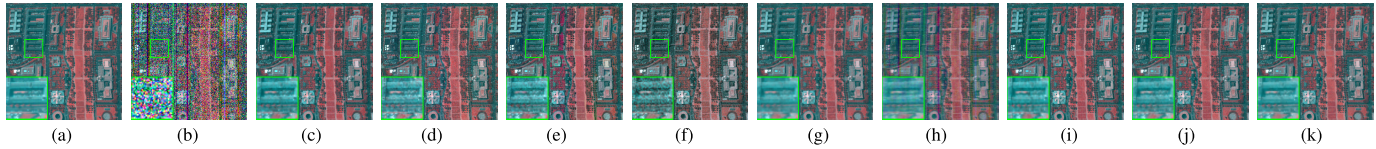
Fig. 1. Denoising results of simulated WDC dataset under Case 5. The false color image is composed by bands (R: 100, G: 9, and B: 7). (a) Original. (b) Noisy. (c) LRTV. (d) LSSTV. (e) WSNM. (f) NMoG. (g) LRTDTV. (h) TLR-SSTV. (i) LRTDGS. (j) FGSLR$_{2/3}$. (k) FGSLR$_{1/2}$.
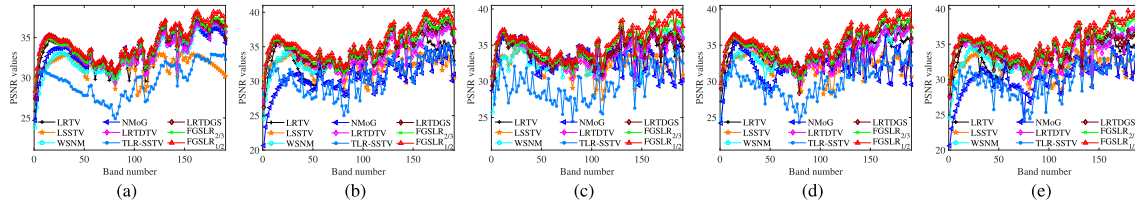

Fig. 2. PSNR values of each band of simulated WDC dataset. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5.
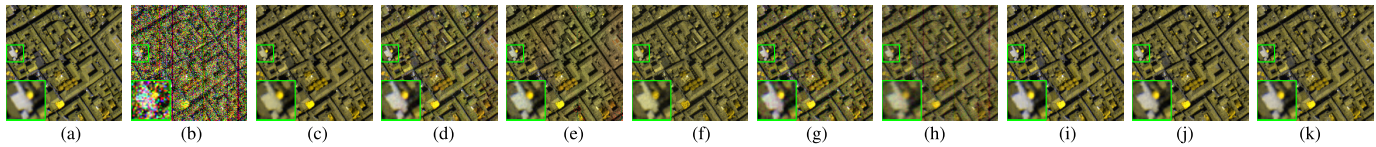

Fig. 3. Denoising results of simulated PaC dataset under Case 5. The false color image is composed by bands (R: 69, G: 57, and B: 5). (a) Original. (b) Noisy. (c) LRTV. (d) LSSTV. (e) WSNM. (f) NMoG. (g) LRTDTV. (h) TLR-SSTV. (i) LRTDGS. (j) FGSLR$_{2/3}$. (k) FGSLR$_{1/2}$.

traditional matrix low-rank denoising methods. Although the PSNR values of LRTDGS are smaller than LRTDTV, the values of another four indices obtained by LRTDGS are more superior to LRTDTV. Summarizing the quantitative indices of the WDC dataset, we can conclude that the proposed FGSLR methods are relatively robustness for HSI denoising since they can achieve the best quantitative results on different datasets under a variety of noise cases.

The visual comparison of the PaC dataset on mixture noise Case 5 is presented in Fig. 3, and we use bands 69, 57, and 5 to

TABLE III

FIVE QUANTITATIVE INDICES COMPARISON OF DIFFERENT HSI DENOISING METHODS ON SIMULATED PaC DATASET

| Case	Index	Noisy	LRTV	LSSTV	WSNM	NMoG	LRTDTV	TLR-SSTV	LRTDGS	FGSLR$_{2/3}$	FGSLR$_{1/2}$
Case 1	PSNR	16.48	31.94	31.82	30.99	31.94	32.59	28.45	32.48	<u>32.67</u>	**32.74**
	SSIM	0.274	0.894	0.898	0.894	0.912	0.913	0.818	0.916	<u>0.923</u>	**0.926**
	FSIM	0.621	0.939	0.942	0.937	0.950	0.947	0.898	0.950	<u>0.952</u>	**0.953**
	ERGAS	553.326	123.127	96.379	105.710	94.479	88.850	138.625	88.356	**86.026**	86.222
	SAM	0.706	0.205	0.156	0.151	0.144	0.145	0.165	0.129	<u>0.115</u>	**0.111**
Case 2	PSNR	13.59	31.85	32.04	31.29	31.84	33.24	29.02	33.21	<u>33.66</u>	**33.85**
	SSIM	0.188	0.900	0.907	0.902	0.901	0.922	0.838	0.928	<u>0.938</u>	**0.941**
	FSIM	0.586	0.939	0.954	0.942	0.948	0.954	0.908	0.959	<u>0.963</u>	**0.964**
	ERGAS	811.341	238.319	93.804	129.431	179.989	88.711	131.418	81.834	<u>77.372</u>	**75.645**
	SAM	0.749	0.295	0.112	0.157	0.187	0.138	0.160	0.106	<u>0.091</u>	**0.090**
Case 3	PSNR	15.01	31.64	32.43	31.27	31.87	33.22	28.63	33.03	<u>33.48</u>	**33.66**
	SSIM	0.238	0.902	0.916	0.910	0.913	0.926	0.832	0.930	<u>0.939</u>	**0.942**
	FSIM	0.626	0.943	0.958	0.946	0.955	0.957	0.905	0.959	<u>0.963</u>	**0.965**
	ERGAS	662.938	156.528	88.546	103.804	108.320	88.863	138.160	82.816	<u>78.034</u>	**77.039**
	SAM	0.696	0.256	0.111	0.126	0.131	0.141	0.172	0.110	<u>0.100</u>	**0.092**
Case 4	PSNR	14.94	32.13	32.58	31.86	32.27	33.39	29.14	33.26	<u>33.88</u>	**33.97**
	SSIM	0.241	0.905	0.914	0.913	0.920	0.923	0.839	0.927	<u>0.940</u>	**0.942**
	FSIM	0.625	0.944	0.958	0.948	0.960	0.955	0.911	0.959	<u>0.963</u>	**0.965**
	ERGAS	674.748	156.824	89.084	97.573	93.066	88.480	130.389	81.623	**75.140**	<u>75.157</u>
	SAM	0.695	0.246	0.106	0.120	0.112	0.136	0.162	0.107	<u>0.096</u>	**0.090**
Case 5	PSNR	13.91	32.09	32.49	31.23	32.47	33.43	29.11	33.48	<u>34.10</u>	**34.30**
	SSIM	0.206	0.905	0.917	0.902	0.910	0.924	0.843	0.933	<u>0.945</u>	**0.947**
	FSIM	0.596	0.943	0.959	0.942	0.952	0.956	0.912	0.962	<u>0.967</u>	**0.968**
	ERGAS	799.014	212.383	88.933	142.491	185.228	91.272	130.965	79.480	<u>74.076</u>	**72.133**
	SAM	0.751	0.302	0.115	0.195	0.218	0.148	0.166	0.105	<u>0.089</u>	**0.087**
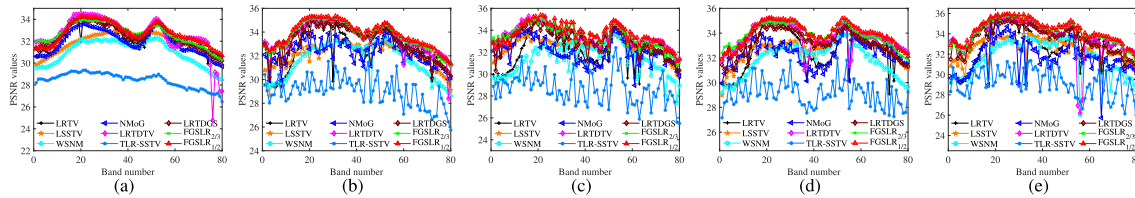
Fig. 4. PSNR values of each band of simulated PaC dataset. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5.

compose the color image for better visualization. Although most of the methods can eliminate the obvious noise, LRTV and NMoG change the contrast as shown in the enlarged box. TLR-SSTV can remove the Gaussian and impulse noise, but the stripe and dead line also exist in the image. Moreover, the Gaussian noise cannot be completely removed by LSSTV and LRTDTV. LRTDGS obtains satisfactory results compared with other methods as displayed in Fig. 3(i). Clearly, the best visual results of FGSLR are shown in Fig. 3(j) and (k), not only removing all mixture noise but also preserving the image details and contrast, whereas other methods achieve poorly.

The values of PSNR of each band on the PaC dataset are presented in Fig. 4. From the results, we can see that LRTV, WSNM, NMoG, LRTDTV, and TLR-SSTV are not very robust since the changing tendency of PSNR on each band are relative instability. Although the PSNR of FGSLR is not the best result in Case 1, we obtain the best result on averaging all bands displayed in Table III. Moreover, FGSLR$_{2/3}$ and FGSLR$_{1/2}$ obtain the obvious superiority in almost all bands under Cases 2–5 when compared with other methods. In summary, the proposed method achieves a significant improvement for HSI denoising in terms of visual and quantitative evaluation.

B. Real Experiments

In this section, we test extensive simulated experiment to illustrate the superiority of our method. To further demonstrate that our method can apply to the real dataset, three real-world HSI are employed for real data test, i.e., GF-5 dataset, HYDICE Urban dataset,[3] and AVIRIS Indian Pines dataset.[4]

1) Experimental Results on GF-5 Dataset: The GF-5 dataset was acquired by the GF-5 satellite, which was developed by the Chinese Aerospace Science and Technology Corporation and launched in 2018. The original size of the GF-5 dataset is $2100 \times 2048 \times 180$, and there are 25 bands that are miss information. After removing the miss bands and extracting a small region, a sub-HSI with the size of $256 \times 256 \times 155$ is chosen for experiments. This dataset is seriously degraded by the mixture of Gaussian, stripes and dead lines noises as shown in Fig. 5(a).

Fig. 5 presents the denoising results on the false color image used bands 96, 154, and 151 to compose. From the results of the enlarged box in Fig. 5, we can easily

[3]http://www.tec.army.mil/hypercube
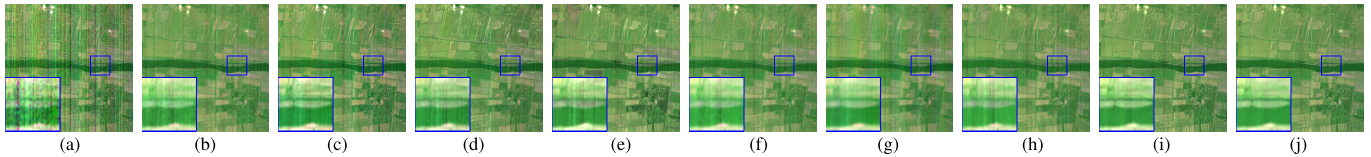[4]https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.html

Fig. 5. Denoising results of real GF-5 dataset. The false color image is composed by bands (R: 96, G: 154, and B: 151). (a) Noisy. (b) LRTV. (c) LSSTV. (d) WSNM. (e) NMoG. (f) LRTDTV. (g) TLR-SSTV. (h) LRTDGS. (i) FGSLR$_{2/3}$. (j) FGSLR$_{1/2}$.
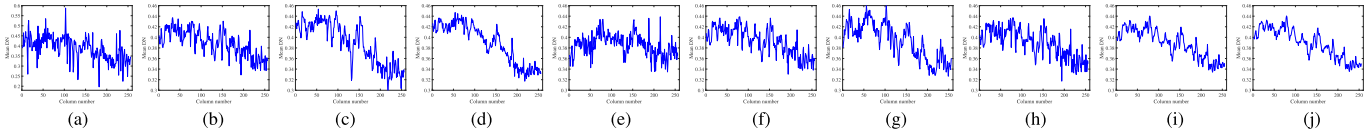

Fig. 6. Vertical mean profiles of real GF-5 dataset on band 154. (a) Noisy. (b) LRTV. (c) LSSTV. (d) WSNM. (e) NMoG. (f) LRTDTV. (g) TLR-SSTV. (h) LRTDGS. (i) FGSLR$_{2/3}$. (j) FGSLR$_{1/2}$.
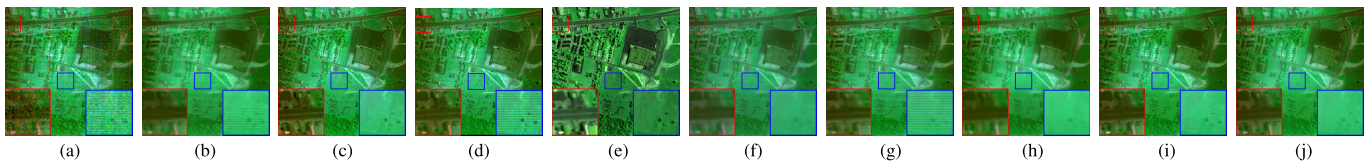

Fig. 7. Denoising results of real Urban dataset. The false color image is composed by bands (R: 207, G: 104, and B: 139). (a) Noisy. (b) LRTV. (c) LSSTV. (d) WSNM. (e) NMoG. (f) LRTDTV. (g) TLR-SSTV. (h) LRTDGS. (i) FGSLR$_{2/3}$. (j) FGSLR$_{1/2}$.
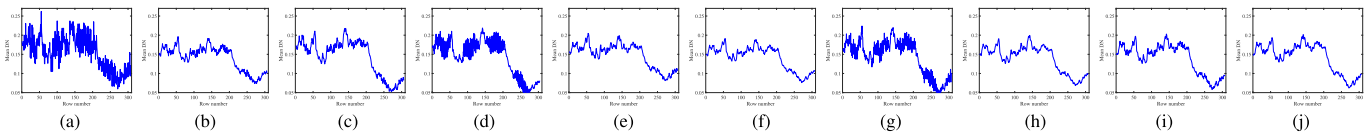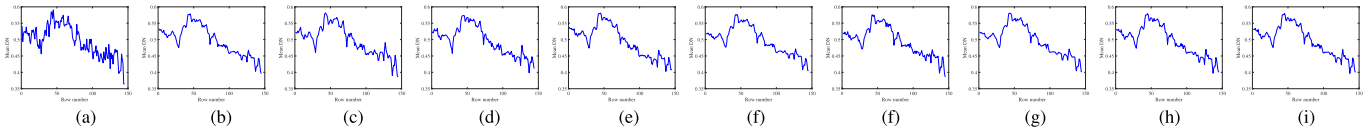

Fig. 8. Horizontal mean profiles of real Urban dataset on band 207. (a) Noisy. (b) LRTV. (c) LSSTV. (d) WSNM. (e) NMoG. (f) LRTDTV. (g) TLR-SSTV. (h) LRTDGS. (i) FGSLR$_{2/3}$. (j) FGSLR$_{1/2}$.

observe that compared methods can remove the Gaussian noise, but fail to remove the stripes and dead lines. As displayed in Fig. 5(i) and (j), the proposed FGSLR methods successfully eliminate the mixture noise as well as restores the underlying image structures. To better understand the existence of stripes in the image, we present the vertical mean profiles of band 154 in Fig. 6. Due to the effects of stripes, there are several fluctuations in the curves of all compared methods. From Fig. 6(i) and (j), the fluctuations are strongly decreased after denoising, which indicates the denoised ability of removing the stripes of FGSLR than other methods.

2) Experimental Results on Urban Dataset: The Urban dataset is a representative real dataset for testing the denoising method due to the complex noise degradation of this data. The size of the original Urban dataset is $307 \times 307 \times 210$, and it is degraded by a variety of noise types, such as Gaussian, stripes, atmosphere, and water absorption. Although some of the bands that are seriously polluted by atmosphere and water cannot provide any information, we also use all bands as the input dataset to confirm the robustness of the denoising methods.

Fig. 7 shows the restored results of false color images composed by bands 220, 150, and 109. As displayed in Fig. 7(a), the image contains many mixture noises. After denoising, the noises are effectively reduced by all methods, but the image details are not handled very well using

some methods. From the enlarged box of Fig. 7(e), NMoG completely removes the noise and restore the image details, but the image contrast is seriously changed. As displayed in the red enlarged box, LSSTV, WSNM, TLR-SSTV, and FGSLR$_{2/3}$ effectively remove the noise and restore the image structure, but there are some residual stripes in another region as shown in the blue enlarged box. LRTDTV and LRTDGS eliminate the noises by blurring the image detail presented in Fig. 7(f) and (h). On the contrary, FGSLR$_{1/2}$ can effectively remove the noise and preserve the spectral since the contrast is not changed. Fig. 8 presents the horizontal mean profiles of the real Urban dataset on band 207. The curves of LRTV, LSSTV, and FGSLR$_{2/3}$ exist minor fluctuations, which indicates the stripes in the image. The blurred results of LRTDTV and LRTDGS can be observed from the smoother horizontal mean profiles produced by them. The curve of FGSLR$_{1/2}$ is more reasonable than other methods as shown in Fig. 8(i).

3) Experimental Results on Indian Pines Dataset: The third real dataset is Indian Pines, which was collected by NASA AVIRIS instrument in Northwestern Indiana in 1992. This dataset contains 145×145 spatial pixels and 220 spectral bands and is mainly polluted by Gaussian, impulse noise, atmosphere, water absorption, and others.

Fig. 9 shows the denoising performance of all compared methods on false color image (R: 220, G: 150, and B: 109).

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

12   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Fig. 9.   Denoising results of real Indian pines dataset. The false color image is composed by bands (R: 220, G: 150, and B: 109). (a) Noisy. (b) LRTV. (c) LSSTV. (d) WSNM. (e) NMoG. (f) LRTDTV. (g) TLR-SSTV. (h) LRTDGS. (i) FGSLR$_{2/3}$. (j) FGSLR$_{1/2}$.

Fig. 10.   Horizontal mean profiles of real Indian pines dataset on band 109. (a) Noisy. (b) LRTV. (c) LSSTV. (d) WSNM. (e) NMoG. (f) LRTDTV. (g) TLR-SSTV. (h) LRTDGS. (i) FGSLR$_{2/3}$. (j) FGSLR$_{1/2}$.
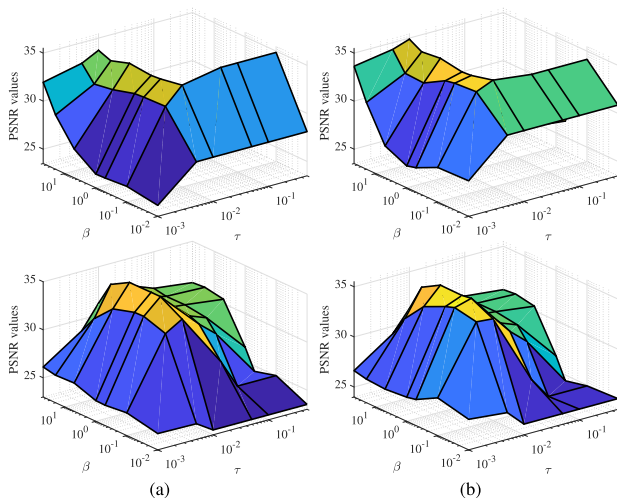

Fig. 11.   Sensitivity analysis of parameters β and τ. (Top row) Case 1. (Bottom row) Case 5. (a) FGSLR$_{2/3}$. (b) FGSLR$_{1/2}$.

Fig. 12.   Sensitivity analysis of parameters λ and μ. (Top row) Case 1. (Bottom row) Case 5. (a) FGSLR$_{2/3}$. (b) FGSLR$_{1/2}$.

From Fig. 9, the image is seriously mixed by Gaussian and impulse noises. LRTV, LRTDTV, TLR-SSTV, and LRTDGS smooth out the image details. LSSTV removes the noise completely, but the spectral information is more or less destroyed as shown in the enlarged box. WSNM obtains the blurry effects in some regions, and NMoG creates artifact information. The superiorly of our proposed FGSLR methods can be reflected by the more natural denoising results compared with other results. This phenomenon also can be observed by the horizontal mean profiles of the blue channel band 109 shown in Fig. 10, FGSLR methods achieve the more suitable results since the curves are not over-smoothed.

C. Discussion of Model Analysis

To further illustrate the effectiveness of the proposed FGSLR methods, the superiorities and robustness of the model parameters should be discussed. There are several parameters in the proposed FGSLR, i.e., regularization parameters β, λ, τ, and α (FGSLR$_{2/3}$ model), penalty parameter μ, proximal parameter ρ, and initial rank d, which are relatively robust to obtain a satisfactory result. From Theorem 1, for any $\alpha > 0$, (8) holds, thus we set $\alpha = 1$ in FGSLR$_{2/3}$ model.
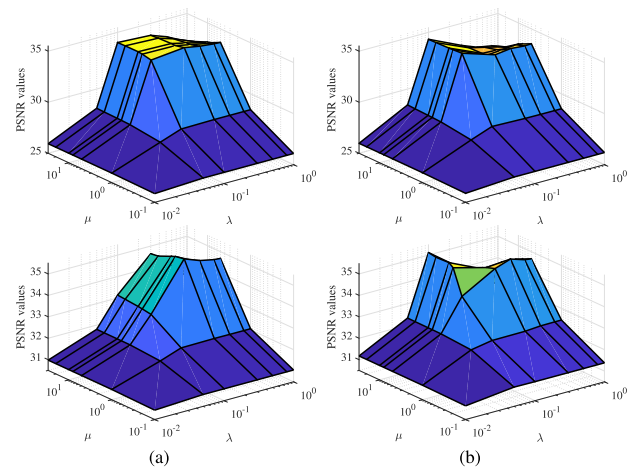
Parameter ρ in proximal term is used to guarantee the theoretical convergence of PAM solver, and $\rho = 0.1$ is set in FGSLR model as recommended in [63]. Moreover, to show the robustness of rank selection of the proposed model, we fix $d = 20$ in all experiments of FGSLR. In the next, we investigate the sensibility of other parameters. Case 1 (Gaussian noise case) and Case 5 (mixed noise case) are selected as experimental object, and we use PSNR as the measure to analyze the change of the parameters.

1) Analysis of β and τ: β and τ are employed to constrain the fidelity term and sparse noise term. Fig. 11 presents the relationship of parameters β and τ. From the results, we can find that the changed curves of PSNR are consistent for FGSLR$_{2/3}$ and FGSLR$_{1/2}$, which indicates the robustness of different FGSLR models with the same parameters. Moreover, the best results are achieved when $\tau = 0.01$ and β is in the range of [0.5, 5]. Based on this observation, we fix $\tau = 0.01$ and $\beta = 0.5$ or 5 for FGSLR methods in all experiments.

2) Analysis of λ and μ: λ is used to constrain the TV term for the factor, and μ is a penalty parameter in the inter ADMM algorithm. The relationship of these two parameters is displayed in Fig. 12. It can be observed that the best results are performed with parameter λ is set 0.5 and 0.1 for

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

CHEN *et al.*: HSI DENOISING USING FGSLR

13

TABLE IV
RUNNING TIME (IN SECONDS) OF DIFFERENT METHODS FOR THE REAL DATASET

dataset	LRTV	LSSTV	WSNM	NMoG	LRTDTV	TLR-SSTV	LRTDGS	FGSLR$_{2/3}$	FGSLR$_{1/2}$
GF-5	144.1	568.2	1384.5	126.9	190.6	3153.8	142.6	158.5	52.0
Urban	451.1	1170.2	4971.9	358.0	741.7	7444.5	429.3	365.3	191.8
Indian Pines	78.8	253.4	1421.9	92.3	116.6	1418.9	96.0	27.2	47.8
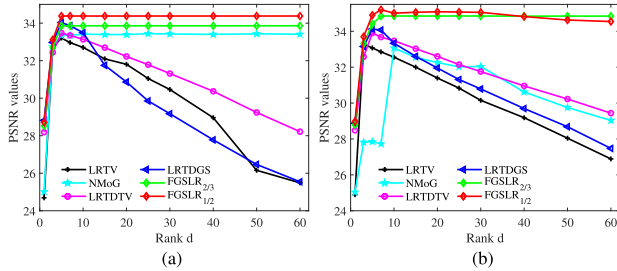
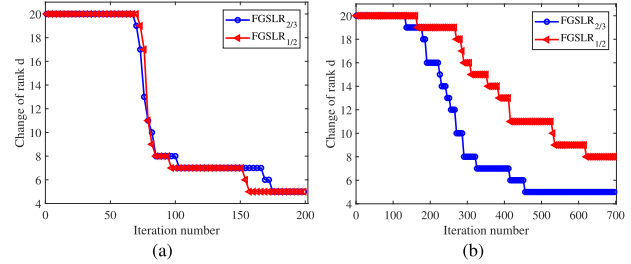Fig. 13. Sensitivity analysis of rank d of different methods. (a) Case 1. (b) Case 5.

Fig. 14. Change of rank d with the iteration increases. (a) Case 1. (b) Case 5.

Fig. 15. Relative change values $(\|\mathcal{B}^{k+1} \times_3 \mathbf{A}^{k+1} - \mathcal{B}^k \times_3 \mathbf{A}^k\|_F)/(\|\mathcal{B}^k \times_3 \mathbf{A}^k\|_F)$ versus the iteration number of the FGSLR solver in the simulated Pavia dataset on Case 5. (a) FGSLR$_{2/3}$. (b) FGSLR$_{1/2}$.

FGSLR$_{2/3}$ and FGSLR$_{1/2}$, respectively. When λ is in the range of $[0.1, 0.5]$, τ is relatively robustness in $[5, 10]$. Therefore, $\lambda = 0.1$ or 0.5 and $\mu = 5$ or 10 are set for FGSLR methods in all experiments.

3) Robustness of Rank Selection: To demonstrate the robustness of rank selection of the proposed FGSLR methods, we compare the sensitivity analysis of rank d for different methods that need to preestablish the rank of HSI. Fig. 13 shows the PSNR curves of different rank d for compared methods. From the results, we can find that LRTV, LRTDTV, and LRTDGS have an optimal rank selection for satisfactory results. When the selected rank is greater than the optimal rank, the results are rapidly decreased, which illustrates the deficiency of this kind of approach. Although the rank selection of NMoG is robust in Case 1, the results are also poor with the increase of rank in Case 5. In contrast, the results of FGSLR methods are very robust with the increase of rank d ($d > r$, r is optimal rank). Therefore, the rank of FGSLR is set $d = 20$ in all experiments, and the compared method is manually adjusted to set the optimal rank.

4) Automatic Adjustment of Rank d: From Fig. 13, we demonstrate the robustness of rank selection of our method. The reason is that we minimize the number of nonzero columns of the factor, and the rank can be automatically adjusted. Fig. 14 presents the change of rank d with the iteration increases. It is obvious to see that the rank d of FGSLR$_{2/3}$ and FGSLR$_{1/2}$ is gradually decreased as the iteration increases and converges to a stable value. Moreover, we can find that the convergence of rank d of FGSLR$_{1/2}$ method is faster than FGSLR$_{2/3}$, which also illustrates the fact that FGSLR$_{1/2}$ obtains the better experimental results and rank approximation than FGSLR$_{2/3}$.

5) Running Time: From the computation complexity of Algorithm 3, FGSLR is a high-efficiency method compared to other low-rank approximation approaches. Table IV lists the running time of different methods for the three real datasets.
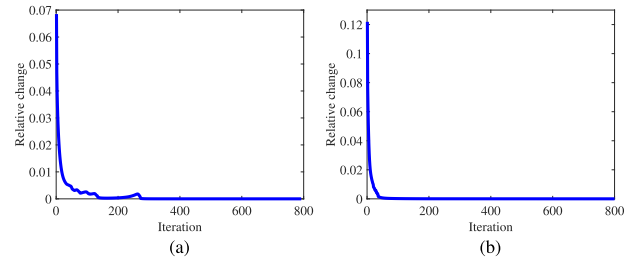
Although the solved algorithm needs to perform the inter and outer iterations, the time cost of the proposed method is fastest compared with other methods, corresponding to the fact of lower computation cost of the FGSLR.

6) Numerical Convergence: Fig. 15 shows the relative change value of the proposed FGSLR methods on the simulated PaC dataset. From the curves, we can see that the relative change value converges to zero as the iteration number increases, indicating the numerical convergence guarantee of the proposed method.

V. CONCLUSION

In this article, two novel FGSLR methods were proposed for HSI denoising. The global spectral correlation of HSI was efficiently depicted by FGSLR, which can embody the advantages of both rank approximations of Schatten-p norm and low-rank factorization, i.e., 1) tighter rank approximation; 2) low computation cost; and 3) robust to rank selection. Moreover, we employed TV regularization to smooth the factor so as to preserve the spatial piecewise constant structure of HSI. A PAM framework with the theoretical convergence guarantee was developed to solve the FGSLR model efficiently. Both simulated and real-word datasets were conducted to demonstrate the performances and superiorities

of the proposed methods than state-of-the-art HSI denoising methods.

In the future, we will try to extend the FGSLR to other types of high-dimension data applications with computation burden or difficulty of rank selection, such as compressive sensing, target detection, and reconstruction of missing information. Moreover, the DCNN can be embedded in FGSLR to fully exploit the spatial information.

ACKNOWLEDGMENT

REFERENCES

[1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[2] D. Hong *et al.*, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[3] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[4] S. Yang, Z. Feng, M. Wang, and K. Zhang, "Self-paced learning-based probability subspace projection for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 630–635, Feb. 2019.

[5] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[6] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[7] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial–spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[8] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.

[9] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[10] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[11] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via non-iterative subspace-based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 7170–7183, 2021.

[12] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.

[13] W. He *et al.*, "Non-local meets global: An integrated paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 29, 2020, doi: 10.1109/TPAMI.2020.3027563.

[14] X.-L. Zhao, W.-H. Xu, T.-X. Jiang, Y. Wang, and M. K. Ng, "Deep plug- and-play prior for low-rank tensor completion," *Neurocomputing*, vol. 400, pp. 137–149, Dec. 2020.

[15] W. Chen, X. Zheng, and X. Lu, "Semisupervised spectral degradation constrained network for spectral super-resolution," *IEEE Geosci. Remote Sens. Lett.*, early access, May 25, 2021, doi: 10.1109/LGRS.2021.3079961.

[16] Y.-T. Wang, X.-L. Zhao, T.-X. Jiang, L.-J. Deng, Y. Chang, and T.-Z. Huang, "Rain streaks removal for single image via kernel-guided convolutional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3664–3676, Aug. 2021.

[17] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a Spatial–Spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Oct. 2019.

[18] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2019.

[19] W. Dong, H. Wang, F. Wu, G. M. Shi, and X. Li, "Deep spatial–spectral representation learning for hyperspectral image denoising," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 635–648, Dec. 2019.

[20] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[21] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[22] Q. Wang, Z. Wu, J. Jin, T. Wang, and Y. Shen, "Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising," *Signal Process.*, vol. 142, pp. 11–26, Jan. 2018.

[23] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, "Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 973–988, Mar. 2019.

[24] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted schatten p-norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.

[25] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *J. Comput. Appl. Math.*, vol. 363, pp. 124–144, Jan. 2020.

[26] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5366–5380, Sep. 2017.

[27] T. Xie, S. Li, and B. Sun, "Hyperspectral images denoising via non-convex regularized low-rank and sparse matrix decomposition," *IEEE Trans. Image Process.*, vol. 29, pp. 44–56, 2020.

[28] L. Zhuang, L. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 8, 2021, doi: 10.1109/TGRS.2020.3046038.

[29] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, early access, Dec. 9, 2021, doi: 10.1109/TGRS.2020.3040221.

[30] F. Xu, Y. Chen, C. Peng, Y. Wang, X. Liu, and G. He, "Denoising of hyperspectral image using low-rank matrix factorization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1141–1145, Jul. 2017.

[31] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. Noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.

[32] Y. Chen, T. Huang, W. He, N. Yokoya, and X. Zhao, "Hyperspectral image compressive sensing reconstruction using subspace-based nonlocal tensor ring decomposition," *IEEE Trans. Image Process.*, vol. 29, pp. 6813–6828, 2020.

[33] T.-X. Jiang, L. Zhuang, T.-Z. Huang, X.-L. Zhao, and J. M. Bioucas-Dias, "Adaptive hyperspectral mixed noise removal," *IEEE Trans. Geosci. Remote Sens.*, early access, Jun. 15, 2021, doi: 10.1109/TGRS.2021.3085779.

[34] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[35] Y. Chen, S. Wang, and Y. Zhou, "Tensor nuclear norm-based low-rank approximation with total variation regularization," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1364–1377, Oct. 2018.

[36] Y.-Y. Liu, X.-L. Zhao, Y.-B. Zheng, T.-H. Ma, and H. Zhang, "Hyperspectral image restoration by tensor fibered rank constrained optimization and plug- and-play regularization," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 5, 2021, doi: 10.1109/TGRS.2020.3045169.

[37] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1692–1700.

[38] Y. Wang, J. Peng, Q. Zhao, D. Meng, Y. Leung, and X.-L. Zhao, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.

[39] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[40] Y. Chang, L. Yan, X. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.

[41] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2019.

[42] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[43] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, "Nonlocal tensor-ring decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1348–1362, Oct. 2020.

[44] L. Zhuang and M. K. Ng, "Hyperspectral mixed noise removal by ℓ_1-norm-based subspace representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1143–1157, 2020.

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

[46] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, and T.-Z. Huang, "Low-rank tensor train for tensor robust principal component analysis," *Appl. Math. Comput.*, vol. 367, p. 124783, 2020.

[47] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 93–107, Nov. 2019.

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

[49] Y. Chang, L. Yan, B. Chen, S. Zhong, and Y. Tian, "Hyperspectral image restoration: Where does the low-rank property exist," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6869–6884, Aug. 2021.

[50] F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient schatten p-norm minimization," in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 655–661.

[51] Y. Zheng, T. Huang, X. Zhao, Y. Chen, and W. He, "Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8450–8464, Dec. 2020.

[52] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, "Framelet representation of tensor nuclear norm for third-order tensor completion," *IEEE Trans. Image Process.*, vol. 29, pp. 7233–7244, 2020.

[53] T.-X. Jiang, X.-L. Zhao, H. Zhang, and M. K. Ng, "Dictionary learning with low-rank coding coefficients for tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 31, 2021, doi: 10.1109/TNNLS.2021.3104837.

[54] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, "Hyperspectral super-resolution via coupled tensor ring factorization," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108280.

[55] J. Fan, L. Ding, Y. Chen, and M. Udell, "Factor group-sparse regularization for efficient low-rank matrix recovery," in *Proc. NIPS*, 2019, pp. 5105–5115.

[56] X. Zheng, Y. Yuan, and X. Lu, "Hyperspectral image denoising by fusing the selected related bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2596–2609, May 2019.

[57] J. Xue, Y. Zhao, W. Liao, J. C. Chan, and S. G. Kong, "Enhanced sparsity prior model for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4567–4581, Nov. 2020.

[58] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 18, 2021, doi: 10.1109/TNNLS.2021.3083931.

[59] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010.

[60] H. Attouch, J. Bolte, and 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.*, vol. 137, nos. 1–2, pp. 91–129, Apr. 2013.

[61] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[62] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, Jan. 2019.

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

Yong Chen received the B.S. degree from the School of Science, East China University of Technology, Nanchang, China, in 2015, and the Ph.D. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020.

From 2018 to 2019, he was a Research Intern with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. He is currently working with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang. His research interests include remote sensing image processing and low-rank matrix/tensor representation.

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.

Prof. Huang is an Editor for *The Scientific World Journal*, *Advances in Numerical Analysis*, *Journal of Applied Mathematics*, *Journal of Pure and Applied Mathematics: Advances in Applied Mathematics*, and *Journal of Electronic Science and Technology*, China.
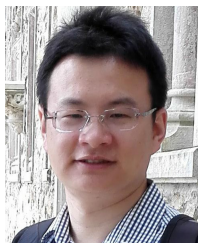
Wei He (Member, IEEE) received the B.S. degree from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, in 2017.

He is currently a Research Scientist with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan, where he was a Researcher, from 2018 to 2020. 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 are model-driven and data-driven methods for image processing problems. His homepage is https://zhaoxile.github.io/.

Hongyan Zhang (Senior Member, IEEE) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

He has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, since 2016. He is a Young Chang-Jiang Scholar appointed by the Ministry of Education of China, Beijing, China. He has authored or coauthored more than 90 research articles and eight talents. His research interests include image reconstruction for quality improvement, hyperspectral information processing and agricultural remote sensing.

Dr. Zhang scored 1st in the Pairwise Semantic Stereo Challenge of the 2019 Data Fusion Contest organized by the IEEE Image Analysis and Data Fusion Technical Committee. He was the Session Chair of the 2015 IEEE WHISPERS conference and the 2016 IEEE IGARSS conference. He serves as an Associate Editor of *Photogrammetric Engineering & Remote Sensing and Computers & Geosciences*. He is a Reviewer of more than 30 international academic journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and so on.

Jinshan Zeng received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2015.

He is currently a Distinguished Professor with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China, and serves as the Director of the Department of Data Science and Big Data. He has authored or coauthored more than 40 articles in high-impact journals and conferences such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), *Journal of Machine Learning Research (JMLR)*, the IEEE TRANSACTIONS ON SIGNAL PROCESSING (TSP), International Conference on Machine Learning (ICML), Association for the Advancement of Artificial Intelligence (AAAI), and so on. His research interests include nonconvex optimization, machine learning (in particular deep learning), and remote sensing.

Dr. Zeng has two articles coauthored with collaborators that received the International Consortium of Chinese Mathematicians (ICCM) Best Paper Award in 2018 and 2020.