Hyperspectral Mixed Noise Removal By ℓ_1-Norm-Based Subspace Representation

Lina Zhuang, *Member, IEEE*, and Michael K. Ng, *Senior Member, IEEE*

***Abstract*—This article introduces a new hyperspectral image (HSI) denoising method that is able to cope with additive mixed noise, i.e., mixture of Gaussian noise, impulse noise, and stripes, which usually corrupt hyperspectral images in the acquisition process. The proposed method fully exploits a compact and sparse HSI representation based on its low-rank and self-similarity characteristics. In order to deal with mixed noise having a complex statistical distribution, we propose to use the robust ℓ_1 data fidelity instead of using the ℓ_2 data fidelity, which is commonly employed for Gaussian noise removal. In a series of experiments with simulated and real datasets, the proposed method competes with state-of-the-art methods, yielding better results for mixed noise removal.**

***Index Terms*—High-dimensional data, hyperspectral destriping, hyperspectral restoration, low-rank representation, nonlocal patch, self-similarity.**

I. Introduction

HYPERSPECTRAL imaging camera measures electromagnetic energy scattered in their instantaneous field view in hundreds or thousands of spectral channels with high spectral resolution. Its high spectral resolution enables materials identification via spectroscopic analysis [1], which leads to countless applications, such as precision agriculture and production monitoring, forest inventories and forest health assessments, water quality assessment, geological mapping, and so on. However, the increase of spectral resolution often implies an increase in the noise corrupted in the image formation process. This degradation mechanism limits the quality of extracted information and its potential applications.

Poissonian noise is becoming the main concern in real hyperspectral imaging [2] as spectral resolution of imagers increases in the new generation hyperspectral sensing systems. The spectral bandwidth decreases implying that, everything else kept constant, each spectral channel receives less photons, yielding higher levels of Poissonian noise. To address Poissonian noise, one way is to solve an optimization problem maximizing the log likelihood based on the Poissonian model [2]. An alternative way is to convert Poissonian noise into approximate additive Gaussian noise with nearly constant variance using variance-stabilizing transformations [3]. This opens a door to use available algorithms conceived for additive Gaussian noise [4], [5]. In this work, we assume that the observation noise is additive Gaussian.

Significant efforts have been made to Gaussian noise removal via sparse representations [6]. Images are sparse and are corresponding to their coefficients with large magnitudes in some transform domains, such as gradient [7], Fourier transform [8], discrete Cosine transform [9], and discrete wavelet transform [10] domains. In contrast, Gaussian noise densely appears in the spatial domain, and it is corresponding to almost zeros transformation coefficients in these low-frequency transform domains. We can shrink the transformation coefficients to attenuate the noise and invert the transform to generate image estimates. Furthermore, it is expected that the sparser the representation is, the more Gaussian noise we can remove [11], [12] from transform domains.

The sparsity of image representation depends on both the way we organize signals [13] and the transform we select. These two aspects are driving the development of machine learning-based denoising methods. On the one hand, self-similarity characteristic of natural images allows us to find similar patches from different locations of images. Grouping similar image patches does enhance the sparsity. In the last decades, nonlocal patch-based methods underlie the state-of-the-art in natural image denoising. We refer to K-SVD [14], BM3D [11], EPLL [15], and WNNM [16]. On the other hand, a proper transform, e.g., data/class-adaptive transform [13], can achieve enhancement of the sparsity of transformation coefficients. In [17, Fig. 3], we show that the subspace basis is data-adaptive and provides a more compact representation than other nonadaptive representations for HSIs. As highly correlated natural images in the spectral-spatial domain, HSIs are also self-similar, low-rank, and piecewise smooth. A number of works [5], [18]–[21], based on low-rank and sparse representations, have been introduced to address Gaussian noise removal.

In practical applications of HSIs, we deal with not only Gaussian noise, but more complicated noise settings, which are linked with various degradation mechanisms. For example, the observed images are corrupted by electronic noise, Poissonian noise, quantization noise, stripe noise, and atmospheric effects. In this article, we focus on the mixture of additive noise, namely Gaussian noise, impulse noise, and stripes.

Since impulse noise and stripes are spatially sparse, a straightforward solution is to model them as an additive sparse term $\mathbf{S}$. In [22]–[25], a ℓ_1 norm is used to induce a randomly sparse

Manuscript received December 4, 2019; revised February 6, 2020; accepted March 1, 2020. Date of publication March 18, 2020. This work was supported in part by the Hong Kong Baptist University Start-up under Grant 21.4551.162562. The work of Ng was supported in part by the Hong Kong Research Grants Council General Research Fund under Grants 12306616, 12200317, 12300218, and 12300519. *(Corresponding author: Lina Zhuang.)*

Lina Zhuang is with the Department of Mathematics, Hong Kong Baptist University (HKBU), Hong Kong (e-mail: linazhuang@hkbu.edu.hk).

Michael K. Ng is with the Department of Mathematics, The University of Hong Kong, Hong Kong (e-mail: mng@maths.hku.hk).

Digital Object Identifier 10.1109/JSTARS.2020.2979801

noise term in their optimization models. The differences between [22]–[25] are the priors used for the clean image. For example, a spatio-spectral total variation (SSTV) of the HSI is minimizing in [22], a weighted nuclear norm of HSI and a TV norm of each band of HSI are jointly minimizing in [23], a group sparse and low-rank tensor decomposition is implemented in [24], a nonconvex regularizer named as normalized ϵ-penalty is used to promote low-rank HSI in [25]. Their different performances are caused by the regularization terms imposing on the clean image. Our work also falls into this line, i.e., a proper prior on the clean image is developed to separate the clean image and the mixed noise. Basically, Gaussian noise can be attenuated effectively by low-rank and sparse representation [5], [18], [19]. Stripes and impulse noise can be removed by total variation and its variants [22], [26].

Apart from ℓ_1 norm, the mixed $\ell_{2,1}$-norm is also used to promote columnwise stripe noise/outlier matrix in [27]–[29]. The works [27] and [28] show that prior of columnwise sparsity is working well for horizontal and vertical stripe artifacts. However, the $\ell_{2,1}$-norm cannot characterize oblique stripes, which are also common in real scenarios. Furthermore, the $\ell_{2,1}$-norm is not suitable for impulse noise, which is spatially random. Alternatively, mixed noise can be addressed by exploiting its statistical distribution, which is approximated by a mixture of Gaussian (MoG) densities in [30].

In addition to the traditional machine learning methods mentioned above, deep learning techniques [31], [32] have revealed in handling mixed noise. A convolutional neural network HSI-DeNet and a spatial-spectral gradient network are presented in [31] and [32], respectively. Promising denoising results are shown, and impressively they are much faster than the traditional machine learning methods. One of the challenges is that with limited training data, the network, for example, HSI-DeNet, is trained for multiband images with a specific noise intensity. However, real remote sensing HSIs have different noise intensity for each band, which cannot be addressed by the HSI-DeNet. If we want to train a network for images with varying noise intensity in the spectral domain, it makes a challenge to the amount of training data.

The motivation of this article is that even though mixed noise is all modeled as a random sparse term in [22]–[25], what leads to different effects is actually their different priors imposing on the clean image. We come to a conclusion that a strong and proper prior of the clean image is crucial for mixed noise removal. The priors of low rankness in spectral domain and self-similarity in spatial domain underlie the state-of-the-art HSI Gaussian denoisers, e.g., FastHyDe [5], GLF [12], and NG-meet [33]. This article exploits the characteristics of low rankness and self-similarity for mixed noise removal. However, stripes also have a low-rank representation. To separate the low-rank clean image and low-rank stripes, we propose a method to estimate a robust subspace, which can represent the clean image well and include less mixed noise.

This article aims at endowing the previous FastHyDe Gaussian denoiser [5] with the ability to address mixed noise. Compared with the article in [5] and other works, the new contributions of this work are as follows.

1) We minimize ℓ_1 norm of reconstruction error instead of adopting ℓ_2 norm because mixed noise is a heavy-tailed distribution instead of a Gaussian distribution. Note that we use the ℓ_1 norm in the data-fidelity term, rather than in the regularization term as other works do [22]–[25].
2) A simple and robust subspace learning is introduced. Subspace representation of HSI is one of the critical reasons for the outstanding performance of work [5]. It enables us to remove the bulk of Gaussian noise and turns the problem into estimating subspace coefficients, which simplifies the estimate. We want to bring the benefit of subspace representation to solve the problem of mixed noise. However, stripes are also low-rank, implying they can be also approximated well in a subspace. If we learn a subspace directly from the noisy data (like the work in [5]), fewer image components and more stripes will remain in the learned subspace (analyzed in Table III).
   To cope with this issue, we propose a new subspace learning method exploiting residual statistics after median filtering, which is very simple but robust to mixed noise.
3) A trick plug-and-play (PnP) prior is used in the optimization model, and the impact of different plugged priors is discussed. The main idea of PnP is that instead of investing efforts in tailoring regularizers promoting self-similar images, we use an off-the-shelf state-of-the-art denoiser directly. In other words, a denoising subproblem in the optimization procedure can be solved by a plugged denoiser (such as BM3D, WNNM, and EPLL). This trick combined with orthogonal transform bridges the gap between hyperspectral denoising and single-band natural image denoising (see more detail in [5, Sec. II-A]). The PnP endows the proposed method with flexibility, meaning any advanced priors of natural images can be plugged and played in our method.

This article is organized as follows. Section II introduces formally L1HyMixDe, a denoising approach conceived for mixed noise. Experimental results comparing with the state of the art are shown in Sections III and IV. Finally, Section V concludes this article.

II. Problem Formulation

A. Observation Model

Let $\mathbf{X} := [\mathbf{x}_1, \dots \mathbf{x}_n] \in \mathbb{R}^{n_b \times n}$ denote a HSI with n spectral vectors (the columns of $\mathbf{X}$) of size n_b. The rows of $\mathbf{X}$ contain n_b spectral bands, which are images, corresponding to the scene reflectance in a given wavelength interval, with n pixels organized in a grid in the spatial domain. In hyperspectral restoration problems under the additive noise assumption, the observation model may be written as

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{n_b \times n}$ represents the observed HSI data and $\mathbf{N} \in \mathbb{R}^{n_b \times n}$ represents additive mixed noise containing Gaussian i.i.d. noise, impulse noise, and stripes.

B. Subspace Representation

Considering the high correlation between channels, we assume that the spectral vectors $\mathbf{x}_i$, for $i = 1, \ldots, n$, lie in a k-dimensional subspace $\mathcal{S}_k$, with $n_b \gg k$. A number of works [1], [5], [34]–[36] have demonstrated that it is a very good approximation in most real HSIs. Therefore, we may write

$$\mathbf{X} = \mathbf{E}\mathbf{Z} \tag{2}$$

where the columns of $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_k] \in \mathbb{R}^{n_b \times k}$ holds a basis for $\mathcal{S}_k$ and matrix $\mathbf{Z} \in \mathbb{R}^{k \times n}$ holds the representation coefficients of $\mathbf{X}$ with respect to (w.r.t.) $\mathbf{E}$. We assume, without loss of generality, that $\mathbf{E}$ is semiunitary, that is $\mathbf{E}^T\mathbf{E} = \mathbf{I}_k$ with $\mathbf{I}_k$ representing the identity matrix of dimension k.

The advantages of using subspace representation include the following three aspects. First, subspace basis $\mathbf{E}$ can be learned from observations (see more details in Section II-C). Second, $\mathbf{Z}$ has a much smaller size than $\mathbf{X}$, implying we are solving a better conditioned problem. Third, comparing to other sparse representations, (orthogonal) subspace representation is a sparser representation (see [17, Fig. 3]), implying signal can be reconstructed well using less coefficients in the transform domain.

C. Subspace Learning Against Mixed Noise

If observations $\mathbf{Y}$ are only contaminated by Gaussian noise, low-dimensional subspace $\mathbf{E}$ is approximately spanned by the first k left-singular vectors of $\mathbf{Y}$, i.e., $\mathbf{E} = \mathbf{U}(:, 1:k)$, where $\mathbf{U} \in \mathbb{R}^{n_b \times n_b}$ is an orthogonal matrix and $\{\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}\} = \text{SVD}(\mathbf{Y})$ with singular values ordered by nonincreasing magnitude. In the case of only Gaussian noise, subspace can be learned directly from the measurements with good approximation [5] because the addition of Gaussian noise increases singular values in each eigenvector direction uniformly and does not change the order of singular values (thus, does not change the estimation of subspace).

In our scenario, however, observations are corrupted not only by Gaussian noise but also by impulse noise and stripes. The subspace of clean image cannot be approximated by the subspace of observations because the addition of impulse noise and stripes increases singular values nonuniformly and changes the order of singular values consequently. Experimental analysis is given in Section III-B.

Herein, we present a robust subspace learning against mixed noise. In order to get a coarse image (denoted as $\widetilde{\mathbf{Y}} = [\tilde{y}_{ij} | 1 \leq i \leq n_b, 1 \leq j \leq n] \in \mathbb{R}^{n_b \times n}$), where Gaussian noise remains, but the bulk of impulse noise and stripes are removed, two image preprocessing steps are conducted as follows.

1) An adaptive median filter is applied to $\mathbf{Y}$ to remove the bulk of impulse noise and stripes as

$$\mathbf{Y}_{\text{med}} = \text{med}(\mathbf{Y}) \tag{3}$$

where $\text{med}(\cdot)$ denotes an adaptive median filtering function [37] implemented band by band, and $\mathbf{Y}_{\text{med}} = [y_{\text{med}ij} | 1 \leq i \leq n_b, 1 \leq j \leq n] \in \mathbb{R}^{n_b \times n}$ is a median filtered image.

2) We can detect pixel indexes of impulse noise and stripes by exploiting the median filtering residual errors $(\mathbf{Y} - \mathbf{Y}_{\text{med}})$. The coarse image $\widetilde{\mathbf{Y}}$ is obtained by replacing the values of impulse noise and stripes with its median filtered values as

$$\tilde{y}_{ij} = \begin{cases} y_{ij} & \text{if } (y_{ij} - y_{\text{med}ij})^2 < \text{th} \\ y_{\text{med}ij} & \text{otherwise} \end{cases} \tag{4}$$

where y_{ij} and $y_{\text{med}ij}$ are elements in the ith row and jth column of the matrix $\mathbf{Y}$ and $\mathbf{Y}_{\text{med}}$, respectively, and the threshold value th is related to the variance of Gaussian noise. If $(y_{ij} - y_{\text{med}ij})^2$ is greater than the threshold, the observation y_{ij} is detected as one affected by impulse noise and stripes. If the observation $\mathbf{y}_{ij}$ is only affected by Gaussian noise, then $(y_{ij} - y_{\text{med}ij})^2$ is close to the variance of Gaussian noise, and we remain its values. Since estimation of intensity of Gaussian noise from observations corrupted by mixed noise is challenging, we suggest that the threshold th can be set using statistic of $(\mathbf{Y} - \mathbf{Y}_{\text{med}})$ as follows:

$$\text{th} = \text{sort}(\mathbf{Y} - \mathbf{Y}_{\text{med}}, p) \tag{5}$$

where function $\text{th} = \text{sort}(\mathbf{X}, p)$ sorts the elements of $\mathbf{X} \in \mathbb{R}^{n_b \times n}$ in the descending order, and returns the element whose sorting index is $\lfloor n_b \times n \times p \rfloor$. In order words, parameter p is the percentage of pixels affected by impulse noise and stripes. In the experiments, we analyze and provide evidence of the robustness of the proposed method w.r.t. overestimation of p.

Given the coarse image $\widetilde{\mathbf{Y}}$, subspace basis can be simply calculated by implementing singular value decomposition (SVD) of $\widetilde{\mathbf{Y}}$ if Gaussian noise in each band is independent and identically distributed (i.i.d.), i.e., the covariance matrix of noise $\mathbf{C}_\lambda = \mathbf{N}\mathbf{N}^T/n = \sigma^2\mathbf{I}$. In practical applications, Gaussian noise in the HSI, however, is band-dependent [1], not i.i.d. Therefore, before implementing SVD of $\widetilde{\mathbf{Y}}$, we whiten the noise in the coarse image and in the observed image, respectively, to have the i.i.d. scenario as

$$\widetilde{\overline{\mathbf{Y}}} = \sqrt{\mathbf{C}_\lambda^{-1}}\widetilde{\mathbf{Y}} \tag{6}$$

and

$$\overline{\mathbf{Y}} = \sqrt{\mathbf{C}_\lambda^{-1}}\mathbf{Y} \tag{7}$$

where $\mathbf{C}_\lambda$ is noise covariance matrix (assumed to be positive definite and estimated by applying the HySime [38] or by any other noise estimation method [39] to $\widetilde{\mathbf{Y}}$), and $\sqrt{\mathbf{C}_\lambda^{-1}}$ denotes the square root of $\mathbf{C}_\lambda^{-1}$ and $\sqrt{\mathbf{C}_\lambda}$ denotes its inverse. For simplicity, we do not change the symbol notations in the left side of (6) and (7). We remark that the Gaussian noise in current observations $\mathbf{Y}$ is i.i.d.

After noise whitening, the signal subspace is approximately spanned by the first k left-singular vectors of $\widetilde{\overline{\mathbf{Y}}}$ as

$$\mathbf{E} = \widetilde{\mathbf{U}}(:, 1:k) \tag{8}$$

where $\widetilde{\mathbf{U}} \in \mathbb{R}^{n_b \times n_b}$ is an orthogonal matrix and $\{\widetilde{\mathbf{U}}, \widetilde{\mathbf{\Sigma}}, \widetilde{\mathbf{V}}\} = \text{SVD}(\widetilde{\overline{\mathbf{Y}}})$ with singular values ordered by nonincreasing magnitude.

D. Cost Function

Given the subspace $\mathbf{E}$, the image restoration problem turns into an estimate of subspace coefficients $\mathbf{Z}$. To deal with mixed noise (namely, Gaussian noise, impulse noise, and stripes), we propose to estimate the matrix $\mathbf{Z}$ by solving the optimization

$$\widehat{\mathbf{Z}} \in \arg\min_{\mathbf{Z}} \|\mathbf{Y} - \mathbf{E}\mathbf{Z}\|_{1,1} + \lambda\phi(\mathbf{Z}) \tag{9}$$

where $\|\mathbf{X}\|_{1,1} := \sum_{i=1}^{n} \|\mathbf{x}_i\|_1$ ($\mathbf{x}_i$ denotes the ith column of $\mathbf{X}$). The regularizer $\phi(\cdot)$ expresses prior information tailored to images. We do not give an implicit formulation of $\phi(\cdot)$ here, because we resort to the PnP [40] prior framework to solve a subproblem related to $\phi(\cdot)$, which is replaced by an off-the-shelf state-of-the-art denoiser.

Because of the ℓ_1 data fidelity, the minimizer in (9) involves an implicit detection of outliers [41]. Observed image elements corrupted only by Gaussian noise are fitted while outlier elements are replaced by estimates determined by the regularization term, independently of the exact value of the outliers. This brings advantage of ℓ_1 data fidelity over the ℓ_2 data fidelity when addressing stripes and impulse noise. We refer to [41] and [42] for the analysis of cost functions composed of an ℓ_1 data-fidelity term and an image regularization term.

We solve the optimization (9) with alternating direction method of multipliers (ADMM) [43] algorithm, which starts by converting the original optimization into a constrained one, using variable splitting, as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{V}} \quad & \|\mathbf{V}\|_{1,1} + \lambda\phi(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{Y} - \mathbf{E}\mathbf{Z} = \mathbf{V}. \end{aligned} \tag{10}$$

The augmented Lagrangian function of the abovementioned optimization is

$$L(\mathbf{Z}, \mathbf{V}, \mathbf{D}) = \|\mathbf{V}\|_{1,1} + \lambda\phi(\mathbf{Z}) + \frac{\mu}{2}\|\mathbf{Y} - \mathbf{E}\mathbf{Z} - \mathbf{V} + \frac{1}{\mu}\mathbf{D}\|_F^2 \tag{11}$$

where $\mu > 0$ is the ADMM penalty parameter.

The application of ADMM to (11) leads to Algorithm 1.[1] The optimization of Line 3 is equivalent to optimization of Line 4 since $\mathbf{E}$ is a semiorthogonal matrix, implying the isometry property $\|\mathbf{E}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ for all $\mathbf{x}$ in $\mathbb{R}^p$.

Lines 4 and 5: these optimizations are the proximity operators (POs) of ϕ applied to $\mathbf{Z}'_t = \mathbf{E}^T(\mathbf{Y} - \mathbf{V}_t + \frac{1}{\mu}\mathbf{D}_t)$ and of the $\ell_{1,1}$-norm applied to $\mathbf{V}'_t = \mathbf{Y} - \mathbf{E}\mathbf{Z}_{t+1} + \frac{1}{\mu}\mathbf{D}_t$, respectively. That is

$$\mathbf{Z}_{t+1} = \Psi_{\lambda\phi/\mu}(\mathbf{Z}'_t) \tag{12}$$

where

$$\Psi_{\lambda\phi}(\mathbf{U}) = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \mathbf{U}\|_F^2 + \lambda\phi(\mathbf{X}) \tag{13}$$

and

$$\mathbf{V}_{t+1} = \Psi_{1\|\cdot\|_{1,1}/\mu}(\mathbf{V}'_t) \tag{14}$$

[1]A MATLAB demo of L1HyMixDe is available in https://github.com/LinaZhuang/L1HyMixDe.git or can be requested by contacting linazhuang@hkbu.edu.hk.

Algorithm 1: ℓ_1-Norm Based **Hy**perspectral **Mix**ed Noise **De**noising (**L1HyMixDe**).

1: Set $t = 0$, choose $\mu > 0$, $\mathbf{V}_0, \mathbf{D}_0$.
2: **repeat**
3: $\mathbf{Z}_{t+1} = \arg\min_{\mathbf{Z}} \lambda\phi(\mathbf{Z}) + \frac{\mu}{2}\|\mathbf{Y} - \mathbf{E}\mathbf{Z} - \mathbf{V}_t + \frac{1}{\mu}\mathbf{D}_t\|_F^2$
4: $= \arg\min_{\mathbf{Z}} \lambda\phi(\mathbf{Z}) + \frac{\mu}{2}\|\mathbf{E}^T(\mathbf{Y} - \mathbf{V}_t + \frac{1}{\mu}\mathbf{D}_t) - \mathbf{Z}\|_F^2$
5: $\mathbf{V}_{t+1} = \arg\min_{\mathbf{V}} \|\mathbf{V}\|_{1,1} + \frac{\mu}{2}\|\mathbf{Y} - \mathbf{E}\mathbf{Z}_{t+1} - \mathbf{V} + \frac{1}{\mu}\mathbf{D}_t\|_F^2$
6: $\mathbf{D}_{t+1} = \mathbf{D}_t + \frac{1}{\mu}(\mathbf{Y} - \mathbf{E}\mathbf{Z}_{t+1} - \mathbf{V}_{t+1})$
7: $t \leftarrow t + 1$
8: **until** stopping criterion is satisfied.

where

$$\Psi_{\lambda\|\cdot\|_{1,1}}(\mathbf{U}) = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \mathbf{U}\|_F^2 + \lambda\|\mathbf{X}\|_{1,1} \tag{15}$$

$$= \text{soft}(\mathbf{U}, \lambda) \tag{16}$$

where $\text{soft}(\cdot, \lambda)$ denotes the componentwise application of the soft-threshold function [44] $x \mapsto \text{sign}(x)\max\{|x| - \lambda, 0\}$.

Optimization (12) is a denoising operator of ϕ. As images in eigen-space (i.e., linear combinations of original bands), rows of $\mathbf{Z}$ have the same characteristic, i.e., self-similarity, as the original image [5]. Self-similarity of rows of $\mathbf{Z}$ is exploited here for solving problem (12). Instead of investing efforts in tailoring regularizers promoting self-similar images, we use directly a state-of-the-art denoiser applied to $\mathbf{Z}'_t$. This is the central idea in PnP [40]. We remark that in the framework of PnP, any off-the-shelf state-of-the-art denoisers conceived to enforce self-similarity, such as BM3D [11], WNNM [16], and EPLL [15], can be adopted for solving subproblem (12). We selected BM3D in this article, as it is the state of the art and its public implementation version is fast.

We do not have a theoretical converge guarantee for the implemented ADMM-PnP since there does not exist a convex regularizer for most state-of-the-art denoisers, such as BM3D and deep learning denoisers. This fact, however, should not prevent us from working in a PnP framework, which allows us to take advantage of state-of-the-art denoisers. The convergence of the PnP iterative procedures is currently an active area of research [40]. In our case, we have systematically observed converge of L1HyMixDe with the augmented Lagrangian parameters set $\mu = 1$ (see numerical convergence in Section III-D).

The computational complexity of L1HyMixDe is dominated by the iterative update of $\mathbf{V}_{t+1}$ and $\mathbf{Z}_{t+1}$. The update of $\mathbf{V}_{t+1}$, i.e., (14), involves only matrix–matrix or matrix–vector multiplications, and thus, its complexity can be easily verified as $\mathcal{O}(knn_b)$. As for the update of $\mathbf{Z}_{t+1}$, it is implemented by a plugged denoiser with input arguments $\mathbf{Z}'_t$, whose computational complexity is $\mathcal{O}(knn_b)$. Assuming the computational complexity of the plugged denoiser is ζ_1,

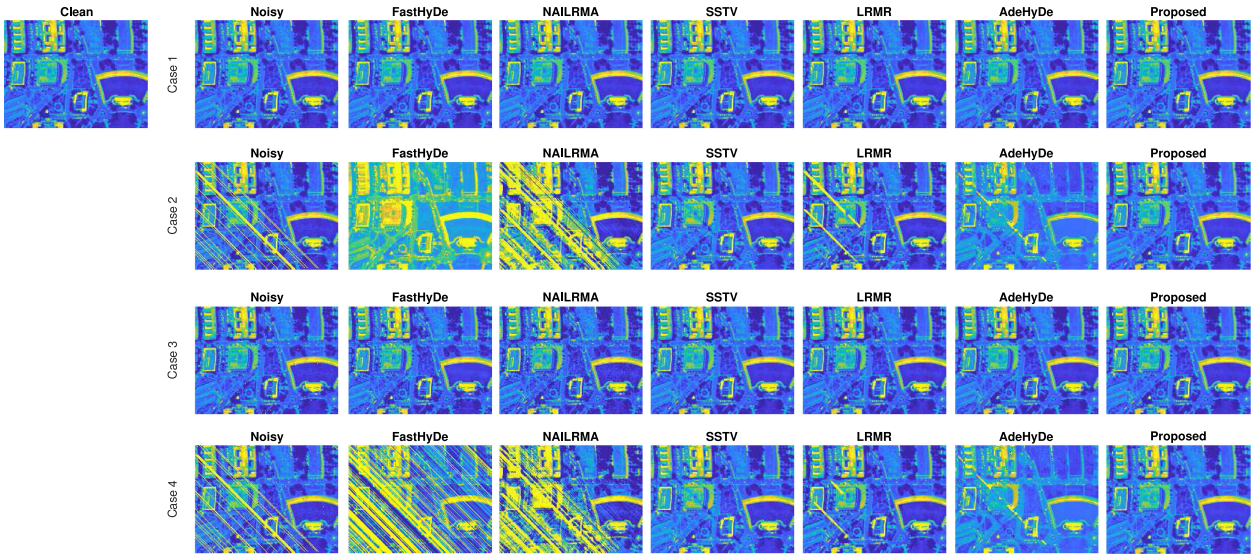


Fig. 1. Denoising band 11 of Washington DC Mall with noise in Case 1 (Gaussian noise), Case 2 (Gaussian noise + stripes), Case 3 (Gaussian noise + "Salt & Pepper" noise), and Case 4 (Gaussian noise + stripes + "Salt & Pepper" noise).

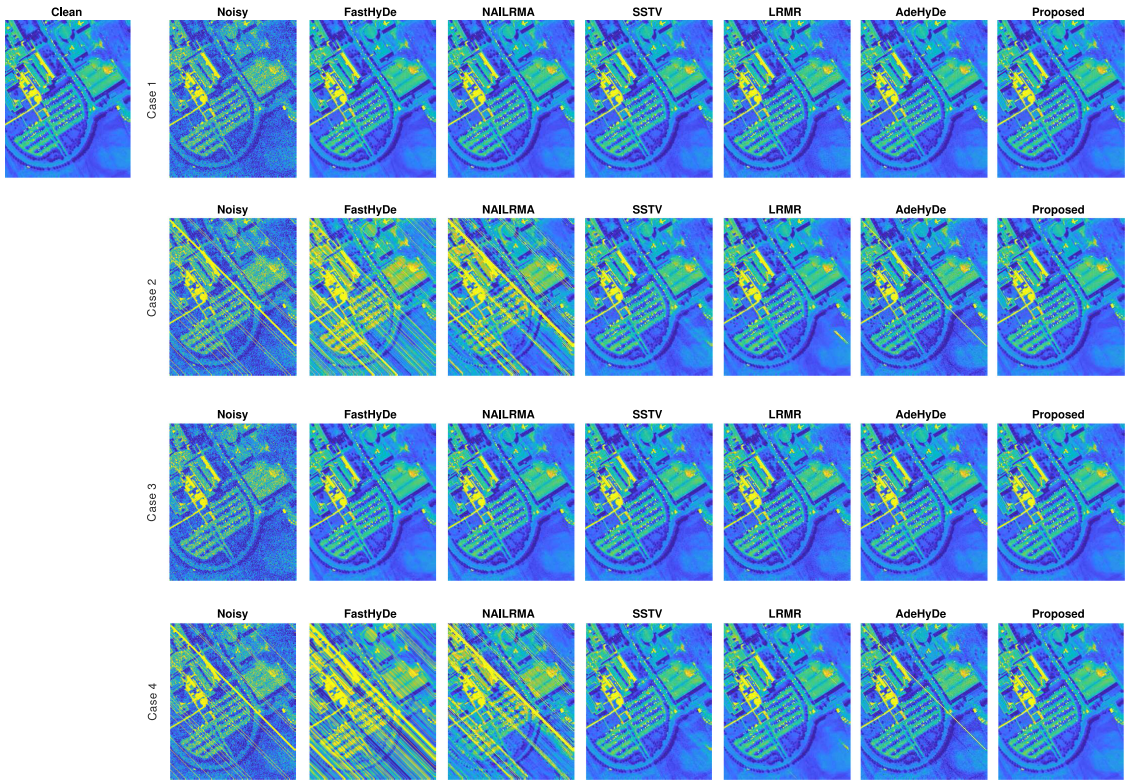


Fig. 2. Denoising band 14 of Pavia University with noise in Case 1 (Gaussian noise), Case 2 (Gaussian noise + stripes), Case 3 (Gaussian noise + "Salt & Pepper" noise), and Case 4 (Gaussian noise + stripes + "Salt & Pepper" noise).

the complexity of computing $\mathbf{Z}_{t+1}$ is $\mathcal{O}(knn_b + \zeta_1)$. Consequently, the overall computational complexity of L1HyMixDe is $\mathcal{O}((knn_b + \zeta_1)\zeta_2)$, where ζ_2 is the number of iterations set in L1HyMixDe.

After solution $\widehat{\mathbf{Z}}$ is obtained, the denoising image can be estimated as follows:

$$\widehat{\mathbf{X}} = \mathbf{E}\widehat{\mathbf{Z}} \tag{17}$$

followed by reconverting noise-whitening space to the original image space as

$$\widehat{\mathbf{X}} = \sqrt{\mathbf{C}_\lambda}\widehat{\mathbf{X}}. \tag{18}$$

III. Experimental Results for Simulated Images

We simulated a noisy hyperspectral dataset (see Fig. 1) as follows [12]. We remove 16 bands with a very low signal-to-noise

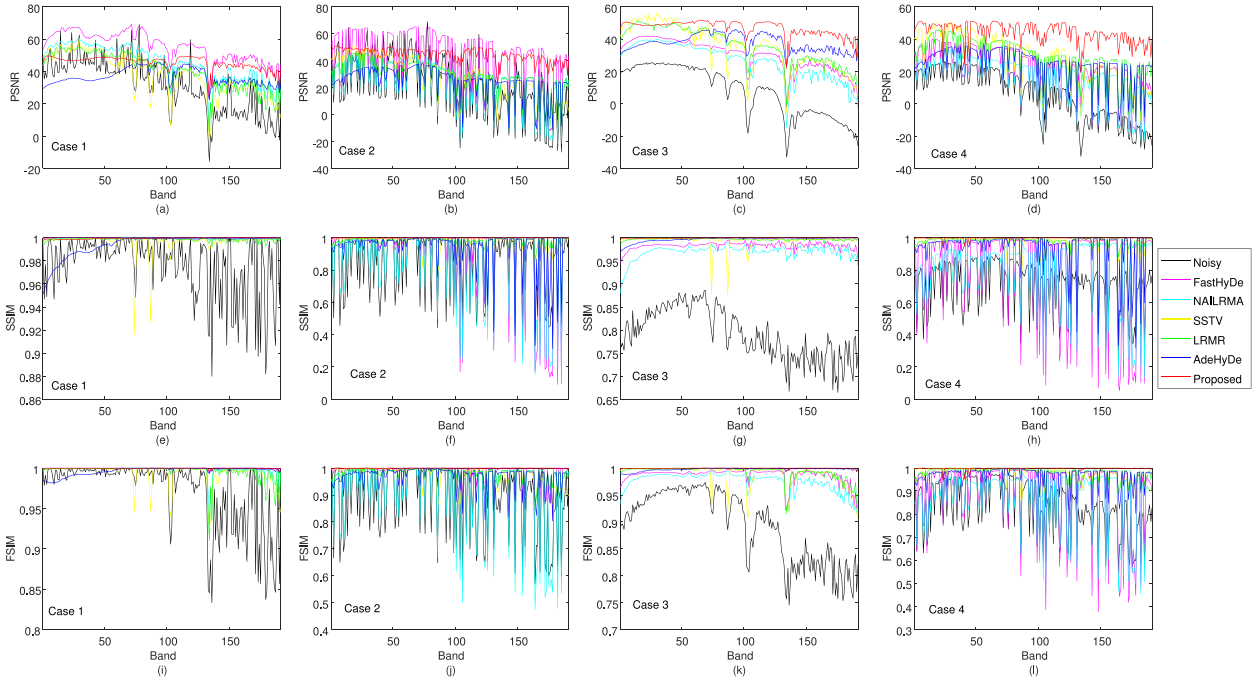


Fig. 3. PSNR (a)–(d), SSIM (e)–(h), and FSIM (i)–(l) values of each band of denoised Washington DC Mall images in Case 1 (Gaussian noise), Case 2 (Gaussian noise + stripes), Case 3 (Gaussian noise + "Salt & Pepper" noise), and Case 4 (Gaussian noise + stripes + "Salt & Pepper" noise).

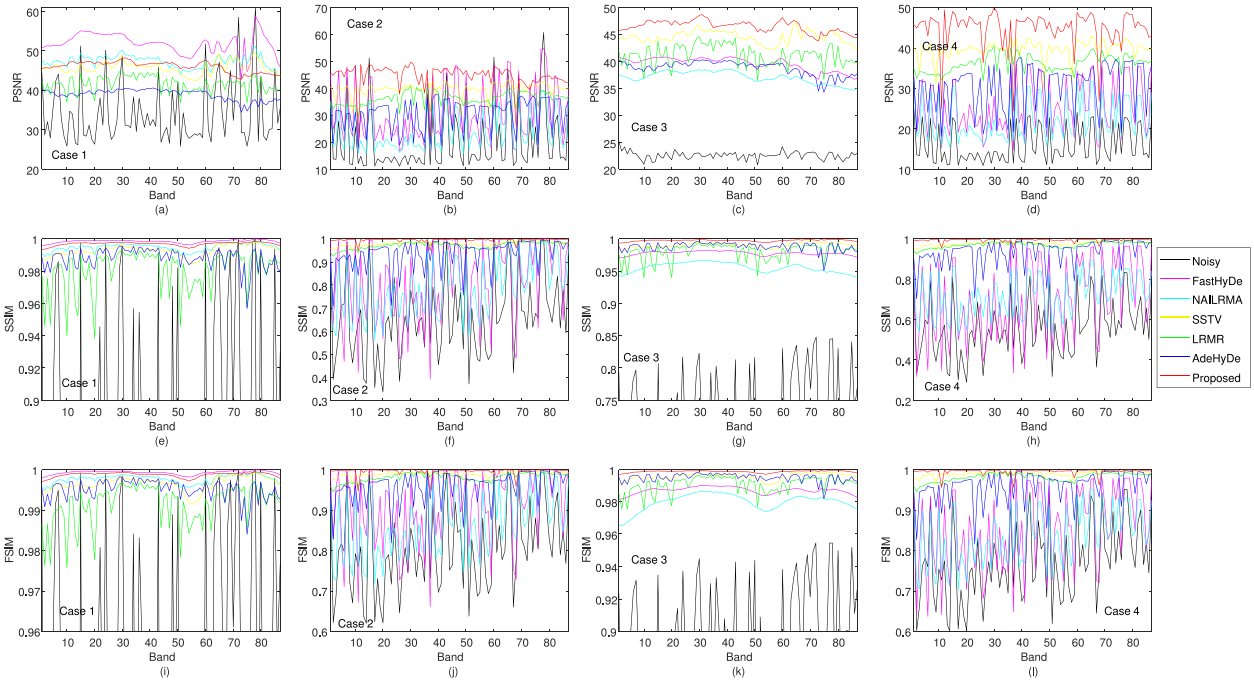


Fig. 4. PSNR (a)–(d), SSIM (e)–(h), and FSIM (i)–(l) values of each band of denoised Pavia University images in Case 1 (Gaussian noise), Case 2 (Gaussian noise + stripes), Case 3 (Gaussian noise + "Salt & Pepper" noise), and Case 4 (Gaussian noise + stripes + "Salt & Pepper" noise).

ratio (SNR) in original Washington DC Mall data[2] considering the signal in these wavelength regions is largely attenuated due to water vapor in the atmosphere leading mostly noise. Then we projected spectral vectors onto the signal subspace of dimension eight learned by implementing SVD on the correlation matrix. The obtained HSI is considered the clean HSI, which was normalized to $[0, 1]$ so that the parameters of competitors can be tuned easier. Four kinds of additive noises are simulated as follows.

[2]The Washington DC Mall data are downloaded from the Purdue University Research Repository (https://engineering.purdue.edu/\,biehl/MultiSpec/hyperspectral.html).

TABLE I
QUANTITATIVE ASSESSMENT OF DIFFERENT DENOISERS APPLIED TO SUBIMAGE OF WASHINGTON DC MALL DATA

		Noisy Image	FastHyDe [5]	NAILRMA [19]	SSTV [22]	LRMR [18]	AdeHyDe [30]	L1HyMixDe
Case 1 Gaussian noise	MPSNR	31.17	**55.50**	45.42	38.17	41.54	36.78	46.82
	MSSIM	0.9762	**0.9999**	0.9995	0.9966	0.9986	0.9944	0.9997
	MFSIM	0.9705	**0.9998**	0.9979	0.9934	0.9957	0.9964	0.9990
	Time (Seconds)	-	**4**	94	164	33	271	120
Case 2 Gaussian noise + stripes	MPSNR	20.80	40.59	24.09	29.52	28.34	24.64	**44.89**
	MSSIM	0.8542	0.8940	0.8559	0.9739	0.9816	0.9366	**0.9994**
	MFSIM	0.8896	0.9836	0.9058	0.9725	0.9784	0.9676	**0.9990**
	Time (Seconds)	-	**4**	76	142	20	252	133
Case 3 Gaussian noise + 'Salt & Pepper' noise	MPSNR	7.92	30.23	26.14	35.24	40.96	37.70	**47.12**
	MSSIM	0.7912	0.9831	0.9694	0.9938	0.9967	0.9969	**0.9997**
	MFSIM	0.8926	0.9865	0.9749	0.9881	0.9922	0.9978	**0.9991**
	Time (Seconds)	-	**3**	55	220	21	265	110
Case 4 Gaussian noise + stripes + 'Salt & Pepper' noise	MPSNR	4.28	20.52	15.98	29.32	28.45	24.42	**44.92**
	MSSIM	0.6953	0.8038	0.7986	0.9734	0.9825	0.9355	**0.9994**
	MFSIM	0.8285	0.8893	0.8678	0.9718	0.9796	0.9675	**0.9991**
	Time (Seconds)	-	**4**	85	159	25	257	113

Note: The bold font indicates the best results.

TABLE II
QUANTITATIVE ASSESSMENT OF DIFFERENT DENOISERS APPLIED TO SUBIMAGE OF PAVIA UNIVERSITY DATA

		Noisy Image	FastHyDe [5]	NAILRMA [19]	SSTV [22]	LRMR [18]	AdeHyDe [30]	L1HyMixDe
Case 1 Gaussian noise	MPSNR	34.29	**51.81**	47.25	45.73	41.59	38.88	48.37
	MSSIM	0.8221	**0.9983**	0.9949	0.9929	0.9807	0.9881	0.9972
	MFSIM	0.9272	**0.9992**	0.9977	0.9966	0.9913	0.9951	0.9987
	Time (Seconds)	-	**7**	93	184	47	841	267
Case 2 Gaussian noise + stripes	MPSNR	20.59	30.97	24.24	40.77	36.35	30.77	**45.12**
	MSSIM	0.6630	0.8107	0.7937	0.9790	0.9725	0.9373	**0.9936**
	MFSIM	0.8068	0.9078	0.8664	0.9901	0.9811	0.9608	**0.9961**
	Time (Seconds)	-	**8**	87	176	38	810	271
Case 3 Gaussian noise + 'Salt & Pepper' noise	MPSNR	22.56	39.56	37.03	44.56	41.26	38.77	**48.37**
	MSSIM	0.6892	0.9780	0.9555	0.9902	0.9795	0.9878	**0.9972**
	MFSIM	0.8780	0.9870	0.9800	0.9957	0.9908	0.9950	**0.9986**
	Time (Seconds)	-	**7**	56	183	52	854	267
Case 4 Gaussian noise + stripes + 'Salt & Pepper' noise	MPSNR	15.97	25.53	22.38	39.84	36.32	30.70	**44.63**
	MSSIM	0.5567	0.7003	0.7241	0.9786	0.9716	0.9371	**0.9927**
	MFSIM	0.7729	0.8526	0.8357	0.9900	0.9807	0.9605	**0.9953**
	Time (Seconds)	-	**8**	97	179	39	811	269

Note: The bold font indicates the best results.

TABLE III
REPRESENTATION ABILITY OF SUBSPACES LEARNED FROM DIFFERENT METHODS

		$\mathbf{E} \leftarrow \text{SVD}(\mathbf{Y})$	$\mathbf{E}$ estimated by RPCA [46]	$\mathbf{E} \leftarrow \text{SVD}(\widetilde{\mathbf{Y}})$ (Proposed)
Case 1 Gaussian noise	γ_1	**1.0000**	**1.0000**	**1.0000**
	γ_2	0.0514	0.0520	**0.0483**
Case 2 Gaussian noise + stripes	γ_1	0.9557	0.9571	**0.9993**
	γ_2	0.3153	0.3138	**0.1675**
Case 3 Gaussian noise + 'Salt & Pepper' noise	γ_1	0.9998	0.9999	**1.0000**
	γ_2	0.0448	0.0440	**0.0409**
Case 4 Gaussian noise + stripes + 'Salt & Pepper' noise	γ_1	0.9558	0.9570	**0.9993**
	γ_2	0.2946	0.2934	**0.1572**

Note: γ_1 and γ_2 denote the relative power of the clean pixels and noise, respectively, that are lying in the estimated subspace, i.e., $\gamma_1 = \frac{\|\mathbf{E}\mathbf{E}^T\mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2}$, and $\gamma_2 = \frac{\|\mathbf{E}\mathbf{E}^T\mathbf{N}\|_F^2}{\|\mathbf{N}\|_F^2}$.
The bold font indicates the best results.

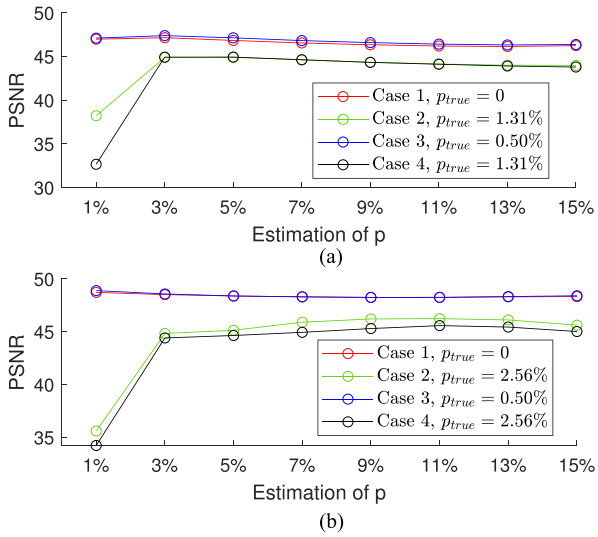


Fig. 5. PSNR values of (a) Washington DC Mall image and (b) Pavia University image denoised by L1HyMixDe as a function of the estimation of p (the percentage of pixels corrupted by impulse noise and stripes).

TABLE IV
IMPACT OF DENOISERS PLUGGED IN THE L1HYMIXDE ON IMAGE RESTORATION IN TERMS OF PSNR (dB)

	Plugged denoisers	BM3D [11]	WNNM [16]	FFDNet [47]	EPLL [15]
Washington DC Mall data					
Case 1	MPSNR	46.82	46.88	42.76	46.90
	Time (Seconds)	120	10352	931	4377
Case 2	MPSNR	44.89	44.99	42.36	44.99
	Time (Seconds)	133	9998	891	4333
Case 3	MPSNR	47.12	47.16	44.22	47.20
	Time (Seconds)	110	10334	901	4264
Case 4	MPSNR	44.92	44.95	43.28	45.01
	Time (Seconds)	113	10453	912	4319
Pavia University data					
Case 1	MPSNR	48.37	48.98	45.13	48.86
	Time (Seconds)	267	29177	1754	12197
Case 2	MPSNR	45.12	45.77	43.58	46.04
	Time (Seconds)	271	27710	1777	12229
Case 3	MPSNR	48.37	49.12	45.15	49.00
	Time (Seconds)	267	26997	1758	12361
Case 4	MPSNR	44.63	45.27	43.24	45.46
	Time (Seconds)	269	27345	1807	12283

Note: Case 1 Gaussian noise, Case 2 Gaussian noise + stripes, Case 3 Gaussian noise + "Salt & Pepper" noise, and Case 4 Gaussian noise + stripes + "Salt & Pepper" noise.

Case 1 (Gaussian non-i.i.d. noise): $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^2)$ where $\mathbf{D}$ is a diagonal matrix with diagonal elements sampled from a uniform distribution $U(0, 0.01)$.

Case 2 (Gaussian noise + stripes): Synthetic data with Gaussian noise (described in Case 1) and oblique stripe noise randomly affecting 30% of the bands and, for each band, about random 10% of the pixels.

Case 3 (Gaussian noise + "Salt & Pepper" noise): Synthetic data with Gaussian noise (described in Case 1) and "Salt & Pepper" noise with noise density 0.5%, meaning affecting approximately 0.5% of elements in $\mathbf{X}$.

Case 4 (Gaussian noise + stripes + "Salt & Pepper" noise): Synthetic data with Gaussian noise (described in Case 1), random oblique stripes (described in Case 2), and "Salt & Pepper" noise (described in Case 3).

Following the abovementioned procedure, we also generate noisy HSIs of Pavia University data[3] (see Fig. 2) subscene.

A. Mixed Noise Removal

We compare the proposed L1HyMixDe with FastHyDe [5], NAILRMA [19], SSTV [22], LRMR [18], and AdeHyDe[4] [30]. All competitors, except for FastHyDe, are conceived to address additive mixed noise in hyperspectral images. FastHyDe is conceived to address pure Gaussian noise and is considered as a benchmark here. FastHyDe, AdeHyDe, and the proposed L1HyMixDe are based on subspace representation. We set the dimension of subspace input to these methods to 8, the true value. For NAILRMA, we set the block size to 20 and step size to 8 because this setting balances the computation and denoising performance. In the paper of NAILRMA [19], the authors use this setting for all testing images. We remark that for methods FastHyDe, NAILRMA, AdeHyDe, and the proposed L1HyMixDe, all of them have a parameter related to Gaussian noise intensity, which is estimated by HySime [38] automatically as their papers do. For SSTV and LRMR, we fine-tuned their parameters of regularizations for all simulated and real images. To save space, we have placed the parameters used at a GitHub link.[5]

The peak signal-to-noise (PSNR) index, the structural similarity (SSIM) index, and the feature similarity (FSIM) index of each band are calculated and depicted in Figs. 3 and 4 for quantitative assessment. Tables I and II give the mean PSNRs (MPSNR), mean SSIMs (MSSIM), and mean FSIM (MFSIM) in Washington DC Mall data and in Pavia Centre, respectively, where we can see for Gaussian noise (in Case 1), the most suitable denoiser is FastHyDe, which, however, is not robust to the mixed noise in Cases 2–4. It is caused by the ℓ_2 data fidelity used in FastHyDe, which squares the error of $(\mathbf{Y} - \mathbf{EZ})$ and is sensitive to outliers (i.e., image elements corrupted by stripes and impulse noise) in the data. To alleviate this, the proposed L1HyMixDe uses ℓ_1 data fidelity in (9) minimizing the sum of the absolute difference between the observations and the estimated image. In Tables I and II, L1HyMixDe yields uniformly the best performance in Cases 2–4 with mixed noise.

As an alternative way to address mixed noise, NAILRMA, SSTV, and LRMR represent mixed noise as an additive sparse matrix (denoted as $\mathbf{S}$) in observation models and use ℓ_2 data fidelity $\|\mathbf{Y} - \mathbf{X} - \mathbf{S}\|_F^2$. When using ℓ_2 data fidelity, a critical issue for this kind of model is that whether the mixture of signal $\mathbf{X}$ and noise $\mathbf{S}$ can be split into two components successfully via proper regularizers on matrix $\mathbf{X}$ and $\mathbf{S}$. Regarding

[3]The Pavia scene is available from the Telecommunications and Remote Sensing Laboratory, Pavia university (Italy) (http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).

[4]We compare with the conference version of AdeHyDe.

[5]https://github.com/LinaZhuang/L1HyMixDe

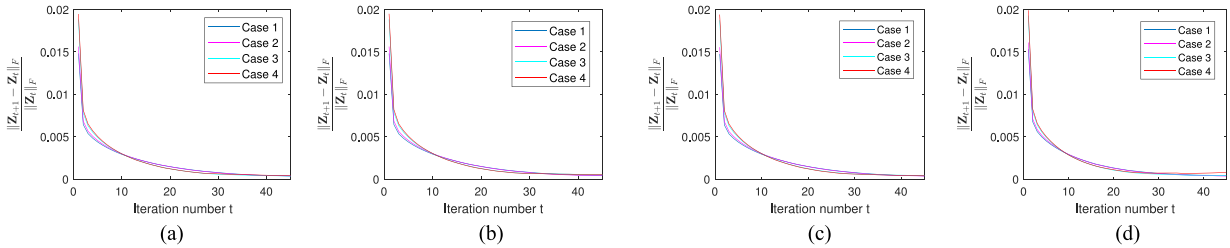


Fig. 6. Relative change of $\mathbf{Z}$ versus the iteration number of L1HyMixDe plugged with different denoisers, namely, (a) BM3D, (b) WNNM, (c) EPLL, and (d) FFDNet when denoising Washington DC Mall datasets.

TABLE V
QUANTITATIVE ASSESSMENT OF DIFFERENT DENOISERS APPLIED TO THE TERRAIN IMAGE AND UNMIXING PERFORMANCE MEASUREMENTS (NMSE_A AND NMSE_S) OF "VCA+FCLS" APPLIED TO THE DENOISED IMAGES

	Noisy Image	FastHyDe [5]	NAILRMA [19]	SSTV [22]	LRMR [18]	AdeHyDe [30]	L1HyMixDe
MPSNR	26.87	46.13	35.65	43.18	39.38	42.06	**51.53**
MSSIM	0.7120	0.9859	0.9540	0.9840	0.9822	0.9925	**0.9980**
MFSIM	0.8581	0.9919	0.9642	0.9900	0.9848	0.9954	**0.9986**
NMSE_A	0.05	0.07	0.06	0.08	0.06	0.59	**0.03**
NMSE_S	0.21	0.27	0.17	0.49	0.18	1.03	**0.15**

Note: The bold font indicates the best results.

Fig. 7. HYDICE data TERRAIN image (R:50, G:36, and B:19).

the regularizer on matrix $\mathbf{X}$, SSTV applies one-dimensional (1-D) total variation on the spectral domain, NAILRMR and LRMR impose spectral low-rankness in spatial patches. These regularizers are not strong enough to preclude severe noise (such as wide stripes) from matrix $\mathbf{X}$. As shown in Figs. 1 and 2, SSTV, LRMR, and AdeHyDe are able to remove light stripes, but leaving some wide stripes. The results of NAILRMA remains heavy noise. On the other hand, L1HyMixDe uses a nonlocal patch-based regularizer on matrix $\mathbf{X}$, promoting self-similar images, which does inject strong prior information.

There are two parameters input to L1HyMixDe: first, the dimension of subspace k, which can be estimated by the HySime [38] method or by any other subspace identification method [45], and second, Parameter of regularization λ, which exists in subproblem (12) and relates to the intensity of Gaussian noise. For example, this article selects BM3D denoiser; thus, λ is corresponding to the variance of Gaussian noise [5], which can be estimated by applying HySime [38] or [39] to the coarse image $\widetilde{\mathbf{Y}}$.

B. Subspace Representation Ability

Subspace representation enables the proposed L1HyMixDe to remove the bulk of Gaussian noise by simply projecting observations onto a subspace. A key issue is whether we can find a proper subspace representing the clean image as much as possible. We compare the representation ability of subspace estimated by three methods, namely SVD of observations, robust principal component analysis (RPCA), and SVD of preprocessed image. Representation ability is measured in terms of the relative power of clean pixels lying in the estimated subspace (denoted as $\gamma_1 = \frac{\|\mathbf{E}\mathbf{E}^T\mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2}$) and relative power of noise lying in the estimated subspace (denoted as $\gamma_2 = \frac{\|\mathbf{E}\mathbf{E}^T\mathbf{N}\|_F^2}{\|\mathbf{N}\|_F^2}$). We can see from Table III, for Case 1 with only Gaussian noise, that all estimated subspaces can represent the clean image fully (i.e., $\gamma_1 = 1.0000$). The proposed subspace estimation method performs slightly better in Case 3 with Gaussian noise and "Salt & Pepper" noise. The advantage of the proposed method is significant in the cases with stripes (Cases 2 and 4), where maximum γ_1 and minimum γ_2 mean more information of the image and less noise remaining in the subspace.

C. Impact of Parameter p

The percentage of pixels corrupted by impulse noise and stripes$p$ is utilized as a threshold value in median filtering residual error $(\mathbf{Y} - \mathbf{Y}_{\text{med}})$ to detect pixels/elements corrupted by impulse noise and stripes. A spectral subspace is learned from observations, where detected pixels/elements are dropped. Therefore, as long as all pixels/elements corrupted by impulse

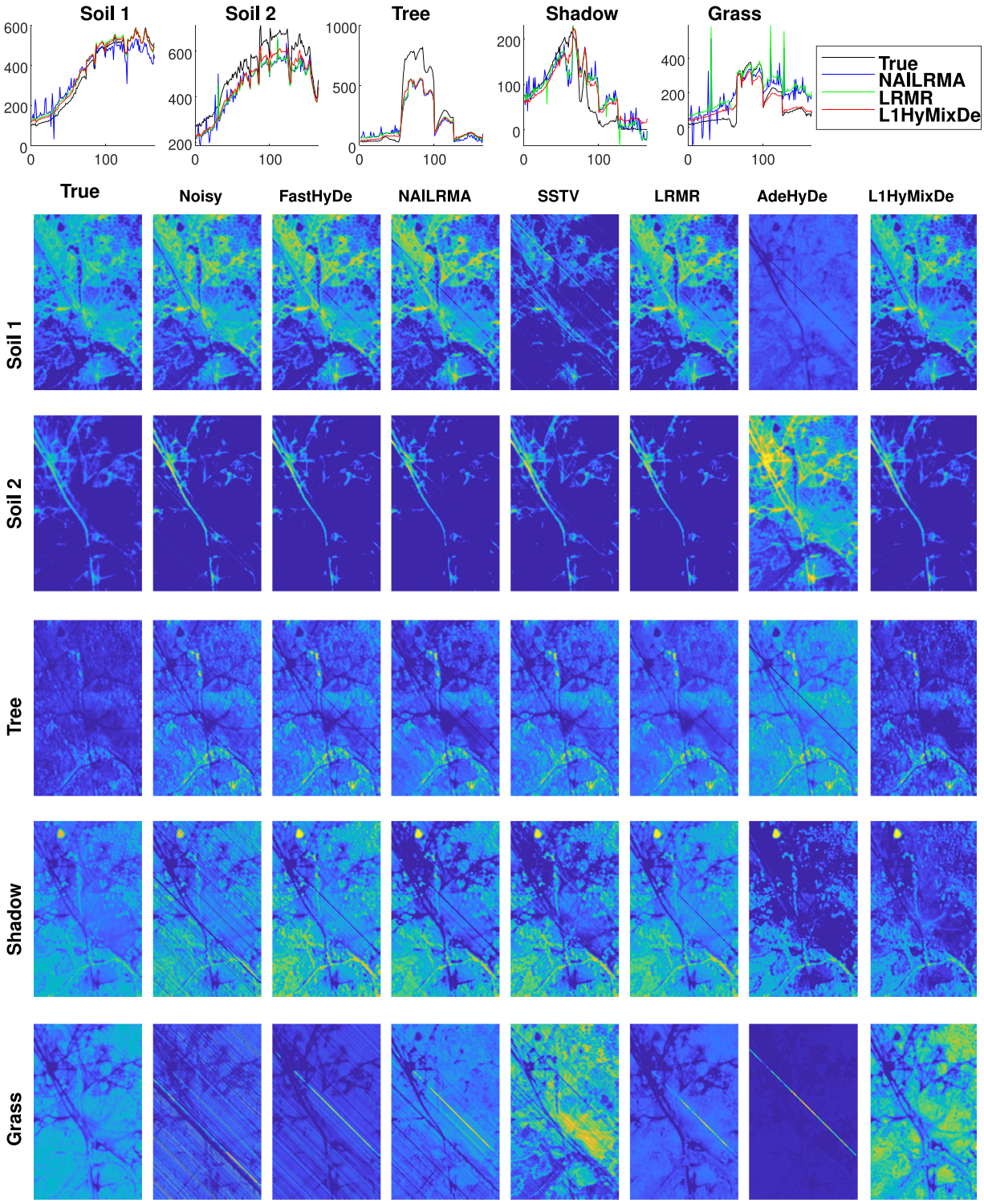


Fig. 8. Estimated endmembers and abundance maps of the TERRAIN images denoised by FastHyDe, NAILRMA, SSTV, LRMR, AdeHyDe, and L1HyMixDe.

noise and stripes are dropped (implying that p can be overestimated), our goal, learning a subspace representing more information of the image and less impulse noise, can be achieved. L1HyMixDe robustness to overestimation of p is illustrated in Fig. 5. Take as example the simulated Washington DC Mall images generated in Cases 2 and 4. The curves in green and black represents the PSNR yielded by L1HyMixDe for different values of p. When p is overestimated (i.e., p $> 1.31\%$, where 1.31% is true value), the corresponding PSNRs are practically constant indicating that L1HyMixDe is very robust to the overestimation of outlier intensity. The identical conclusion can also be drawn from other PSNR curves in Fig. 5. Because of the L1HyMixDe robustness to overestimation of p, we simply fix $p = 5\%$ for all experiments in this article.

D. Impact of Plugged Denoiser

The proposed L1HyMixDe is flexible in the sense that any off-the-shelf denoisers can be adopted to solve the subproblem (12). We take four state-of-the-art denoisers for instance to discuss the impact of plugged denoiser. Table IV gives PSNRs of images restored by L1HyMixDe plugged with three traditional machine learning-based methods (BM3D [11], WNNM [16], and EPLL [15]), and one deep learning method FFDNet [47]). The first three methods show comparable results, better than those of FFDNet. The main reason leading to the difference might be the patch size used. The available network of FFDNet was trained using image patches of size 70×70, whereas other methods all set patch size to 8×8. Considering hyperspectral remote sensing images have a relatively low spatial resolution, the patch size (70×70) may be too large. Furthermore, comparing the results in Tables I and II, the counterparts of L1HyMixDe plugged with different denoisers (WNNM, EPLL, and FFDNet) in Table IV are competitive. From the view point of PSNR, denoisers BM3D, WNNM, EPLL, and FFDNet are equally good for addressing pure Gaussian noise. But if we plug the denoiser in an iterative procedure, we concern with its computation complexity. BM3D may be preferred as its public implementation version is fast (see running times in Table IV).

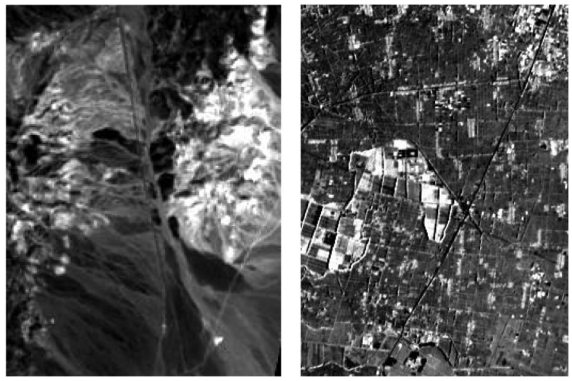
E. Numerical Convergence of the L1HyMixDe

If the plugged denoiser does not have a convex regularizer (as it is for most state-of-the-art denoisers), then the L1HyMixDe is a nonconvex optimization problem, which is hard to prove the convergence of the algorithm. Here, we give an empirical analysis for the convergence of the solver ADMM-PnP with different denoisers. We set the augmented Lagrangian parameters as $\mu = 1$. Curves of the relative change $\frac{\|\mathbf{Z}_{t+1}-\mathbf{Z}_t\|_F}{\|\mathbf{Z}_t\|_F}$ values versus the iteration number are given in Fig. 6, where we can observe that the relative change converges to zero after 40 iterations, implying the convergence of the L1HyMixDe plugged with these four denoisers can be numerically guaranteed.

F. Application in Hyperspectral Unmixing

As image denoising is usually a preprocessing step for successive applications, we take hyperspectral unmixing as an example and conduct experiments to show the impact of denoising on hyperspectral unmixing. We simulate a clean semireal HSI based on the public available TERRAIN image (see Fig. 7) following the generation steps in [48].[6] The original TERRAIN image has size 500 (rows) $\times$ 307 (columns) $\times$ 166 (bands), and is mainly composed of soil, tree, grass, and shadows. The number of endmembers is empirically set to 5 like [48]–[50]. Briefly, a clean TERRAIN image is synthesized based on the linear mixing model, i.e., $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A}$ and $\mathbf{S}$ are the matrices of endmember and abundance, respectively, estimated from the original TERRAIN image. Next, we generate a noisy TERRAIN HSI by adding the Gaussian noise and oblique stripe noise (as described in Case 2), i.e., $\mathbf{Y} = \mathbf{AS} + \mathbf{N}$, yielding MPSNR $=$ 26.87 dB.

[6]A MATLAB demo describing image simulation can be found in https://github.com/LinaZhuang/NMF-QMV_demo.



(a) (b)

Fig. 9. Real HSI used in the experiments. (a) Subimage of Hyperion Cuprite image (Band: 30). (b) Subimage of Tiangong-1 sensor (Band: 3).

Denoisers are applied to the noisy TERRAIN image. A quantitative assessment of denoised images is given in Table V. We implement spectral unmixing of the noisy and denoised images with vertex component analysis (VCA) estimating endmembers and fully constrained least-squares (FCLS) estimating abundances. Unmixing performance is measured via normalized mean square error (NMSE) of endmembers $\mathbf{A}$ and abundances $\mathbf{S}$, denoted as NMSE_A and NMSE_S, respectively. We can see from Table V the image denoising, implemented by NAILRMA, LRMR, and the propose L1HyMixDe, does improve the performance of spectral unmixing. Other denoisers (namely, FastHyDe, SSTV, AdeHyDe) may give good denoising results in terms of global measurement indices (namely, MPSNR, MSSIM, and MFSIM); however, very few stripe noise still remains in the denoising results and affects the estimation of Grass endmember and its abundance map (see the final row of Fig. 8). If we only focus on the endmember of Grass, the one estimated from the denoised image of L1HyMixDe is closest to the true one (see the first row of Fig. 8), also leading to the best abundance estimation of Grass.

IV. Experimental Results for Real Images

A. Hyperion Cuprite Dataset

In this section, we apply denoisers to a Hyperion HSI dataset [in Fig. 9(a)]. It was captured at Cuprite, NV, USA, by Hyperion sensor, which has 242 channels with a spectral resolution of 10 nm, covering from 355 to 2577 nm. It has a spatial resolution of 30 m/pixel. The subimage used in this article has 240×178 pixels with 177 spectral channels (after removing very low SNR channels). Fig. 10 shows four noisy bands in the first columns.

For FastHyDe, AdeHyDe, and L1HyMixDe methods, we empirically set the dimension of subspace to 7. The parameter setting of all competitors can be found at a Github link.[5]

The denoising results and computational times are shown in Fig. 10 and the figure caption, respectively. The noisy image

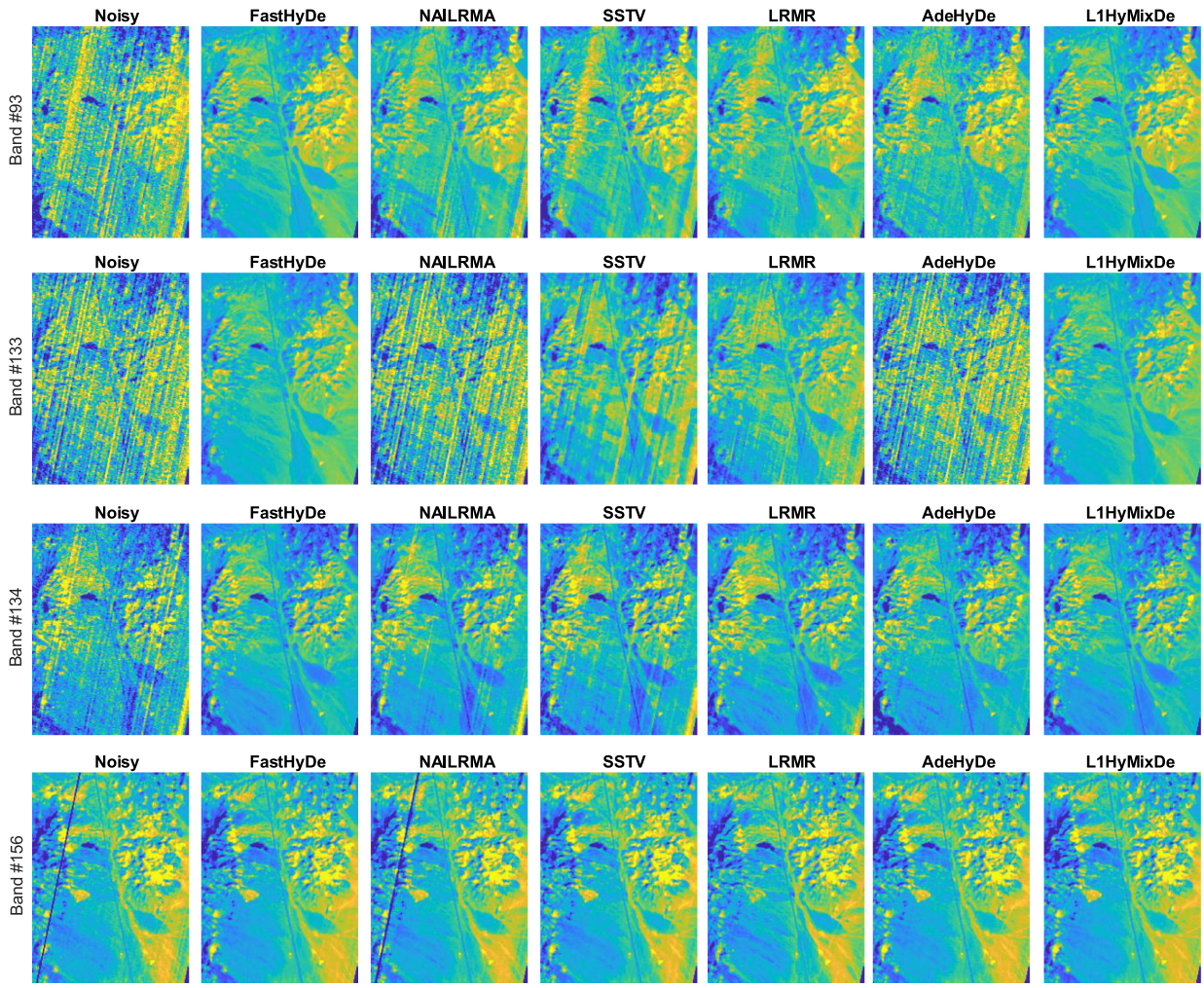


Fig. 10. Denoising results of FastHyDe (4 s), NAILRMA (108 s), SSTV (201 s), LRMR (26 s), AdeHyDe (41 s), and the proposed L1HyMixDe (37 s) in Hyperion Cuprite dataset.

has obvious stripes directing from the upper right to the lower left. The stripes appear to be spatially random, meaning they do not exist in specific locations over all bands. For this kind of stripes and noise, subspace representation is a powerful tool. For example, if we focus on the lower-right corner of 34th band in Fig. 10, an obvious wide stripe is removed by FastHyDe, AdeHyDe, and L1HyMixDe, which are all based on subspace representation. This wide stripe exists in only few bands, implying it lies outside the signal subspace. Therefore, the high correlation between bands can be exploited to remove it. The rationale of proposed L1HyMixDe is similar to that of AdeHyDe, both taking advantage of spectral low-rankness and spatial self-similarity. AdeHyDe models Gaussian noise and sparse noise as a MoG mixture, where stripes and impulse noise is assumed to be a zero-mean Gaussian distribution with a very large variance. Whether the MoG model works depends on the distribution of real sparse noise. L1HyMixDe avoids modeling the complicated noise, and it still works well in the Hyperion Cuprite dataset.

B. *Tiangong-1 Dataset*

A HSI data acquired by one of the sensors in Tiangong-1 (TG-1) imager are used in this section. The sensor is a 75-band push broom scanner with nominal bandwidths of 23 nm short wave infrared, covering from 800 to 2500 nm. The test subimage, recorded over an area of Qinghai Province, China in May 2013, has 351×253 pixels [in Fig. 9(b)]. Four noisy bands (1, 8, 31, and 59) are shown in the first column of Fig. 11.

For FastHyDe, AdeHyDe, and L1HyMixDe methods, we empirically set the dimension of subspace to 3, because we visually found that its fourth band of the projected image has very low SNR. The parameter setting of all competitors can be found at a Github link.[5]

Fig. 11 shows the denoising results of four bands. Instead of being spatially random in Hyperion Cuprite data, stripes in this Tiangong-1 data exist in specific locations over bands, which is a more challenging case. Pixels affected by stripes are noisy over bands, implying spatial information rather than spectral information is critical for destriping. Spatial self-similarity, used

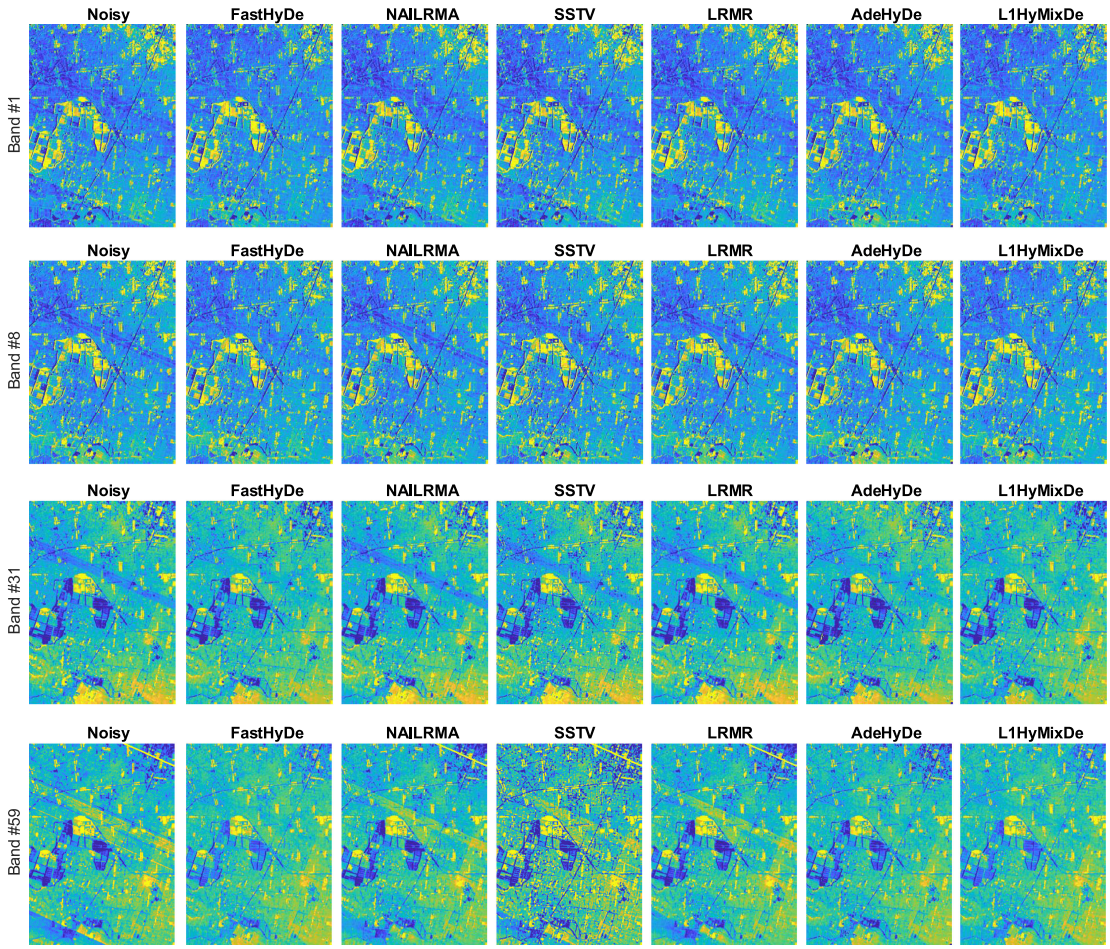


Fig. 11. Denoising results of FastHyDe (4 s), NAILRMA (67 s), SSTV (134 s), LRMR (31 s), AdeHyDe (347 s), and the proposed L1HyMixDe (46 s) in Tiangong-1 dataset.

in FastHyDe, AdeHyDe, and L1HyMixDe, is a stronger prior than total variation used in SSTV and patchwise low-rankness used in NAILRMA and LRMR. In Fig. 11, L1HyMixDe visually yields better results, while FastHyDe and AdeHyDe still remain slight stripes in 31st band.

V. Conclusion

This article proposes a new HSI denoising method, termed L1HyMixDe, addressing the MoG noise, impulse noise, and stripes. The results demonstrate it is possible to sidestep the model of mixed noise. Instead of investing efforts in conceiving a universal regularization for various additive noise, we replace ℓ_2 data fidelity by ℓ_1 data fidelity, which is robust to impulse noise and stripes. It works since we impose strong prios/regularizers on the clean image, namely, low-rankness in the spectral domain and self-similarity in the spatial domain. Considering stripe noise is also low-rank, a subspace directly learned from observations can represent both image components and stripes well. The new subspace estimation method we proposed is able to represent the clean image well and include less stripe noise, leading to a better denoising performance. Some state-of-the-art single-band denoisers (namely, BM3D, WNNM, EPLL, and FFDNet) promoting self-similar image have been used in a PnP manner. Their promising denoising results in the experiments show that any advanced priors of nature image can be plugged and played in our framework. A comparison of L1HyMixDe with the state-of-the-art algorithms is conducted, leading to the conclusion that L1HyMixDe yields better performance for mixed noise.

References

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

[2] L. He, J. A. O'Sullivan, D. V. Politte, K. E. Powell, and R. E. Arvidson, "Quantitative reconstruction and denoising method hyber for hyperspectral image data and its application to crism," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1219–1230, Apr. 2019.

[3] M. Makitalo and A. Foi, "A closed-form approximation of the exact unbiased inverse of the anscombe variance-stabilizing transformation," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2697–2698, Sep. 2011.

[4] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[5] L. Zhuang and J. 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.

[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] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Scientific Comput.*, vol. 20, no. 6, pp. 1964–1977, 1999.

[8] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.

[9] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.

[10] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.

[11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[12] L. Zhuang and J. Bioucas-Dias, "Hyperspectral image denoising based on global and non-local low-rank factorizations," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 1900–1904.

[13] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral–spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan. 2016.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[15] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.

[16] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2862–2869.

[17] L. Zhuang and J. Bioucas-Dias, "Hy-demosaicing: Hyperspectral blind reconstruction from spectral subsampling," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 4015–4018.

[18] 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.

[19] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[20] Y. Chang, L. Yan, and S. Zhong, "Hyper-Laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4260–4268.

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

[22] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.

[23] B. Du, Z. Huang, N. Wang, Y. Zhang, and X. Jia, "Joint weighted nuclear norm and total variation regularization for hyperspectral image denoising," *Int. J. Remote Sens.*, vol. 39, no. 2, pp. 334–355, 2018.

[24] Z. Huang, S. Li, L. Fang, H. Li, and J. A. Benediktsson, "Hyperspectral image denoising with group sparse and low-rank tensor decomposition," *IEEE Access*, vol. 6, pp. 1380–1390, 2018.

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

[26] Y. Chang, L. Yan, H. Fang, and H. Liu, "Simultaneous destriping and denoising for remote sensing images with unidirectional total variation and sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1051–1055, Jun. 2014.

[27] Y. Chen, T.-Z. Huang, and X.-L. Zhao, "Destriping of multispectral remote sensing image using low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4950–4967, Dec. 2018.

[28] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

[29] L. Zhuang, L. Gao, B. Zhang, and J. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," in *Proc. SPIE Remote Sensing XXIII*, Oct. 2017, vol. 10427, Art. no. 104270M.

[30] T. Jiang, L. Zhuang, T. Huang, and J. M. Bioucas-Dias, "Adaptive hyperspectral mixed noise removal," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 4035–4038.

[31] 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.

[32] 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.

[33] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6868–6877.

[34] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[35] R. Dian, S. Li, L. Fang, and J. Bioucas-Dias, "Hyperspectral image super-resolution via local low-rank and sparse representations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 4003–4006.

[36] L. Gao *et al.*, "Subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 349–353, Feb. 2015.

[37] R. C. Gonzalez, S. L. Eddins, and R. E. Woods, *Digital Image Publishing Using MATLAB*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004.

[38] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[39] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 488–498, Apr. 2013.

[40] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

[41] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imag. Vis.*, vol. 20, no. 1-2, pp. 99–120, 2004.

[42] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms application to the processing of outliers," *SIAM J. Numer. Anal.*, vol. 40, no. 3, pp. 965–994, 2002.

[43] J. Eckstein and Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992.

[44] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[45] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

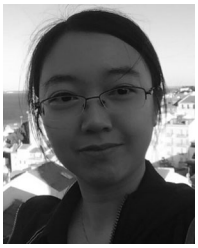
[46] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.

[47] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[48] L. Zhuang, C. Lin, M. A. T. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9858–9877, Dec. 2019.

[49] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[50] Y. Altmann, M. Pereyra, and J. M. Bioucas-Dias, "Collaborative sparse regression using spatially correlated supports-application to hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5800–5811, Dec. 2015.



Lina Zhuang (Member, IEEE) received the bachelor's degrees in geographic information system and in economics from South China Normal University, Ghuangzhou, China, in 2012, the M.S. degree in cartography and geography information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2015, and the Ph.D. degree in electrical and computer engineering from the Instituto Superior Tecnico, Universidade de Lisboa, Lisbon, Portugal, in 2018.

Since 2015, she has been with the Instituto de Telecomunicações, as a Marie Curie Early Stage Researcher of Sparse Representations and Compressed Sensing Training Network (SpaRTaN number 607290). SpaRTaN Initial Training Networks (ITN) is funded under the European Union's Seventh Framework Programme (FP7-PEOPLE-2013-ITN) call and is part of the Marie Curie Actions–ITN Funding Scheme. Currently, she is a Research Assistant Professor with the Hong Kong Baptist University, Hong Kong. Her research interests include hyperspectral image restoration, superresolution, and compressive sensing.

Michael K. Ng (Senior Member, IEEE) received the B.Sc. and M.Phil. degrees from the University of Hong Kong, Hong Kong, in 1990 and 1992, respectively, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 1995.

He is currently a Chair Professor with the Department of Mathematics, the University of Hong Kong. He serves on the editorial boards of international journals. His research interests include bioinformatics, image processing, scientific computing, and data mining.