

# Denoising Hyperspectral Image With Non-i.i.d. Noise Structure

Yang Chen, Xiangyong Cao, Qian Zhao, Deyu Meng, *Member, IEEE*, and Zongben Xu

**Abstract**—Hyperspectral image (HSI) denoising has been attracting much research attention in remote sensing area due to its importance in improving the HSI qualities. The existing HSI denoising methods mainly focus on specific spectral and spatial prior knowledge in HSIs, and share a common underlying assumption that the embedded noise in HSI is independent and identically distributed (i.i.d.). In real scenarios, however, the noise existed in a natural HSI is always with much more complicated non-i.i.d. statistical structures and the under-estimation to this noise complexity often tends to evidently degenerate the robustness of current methods. To alleviate this issue, this paper attempts the first effort to model the HSI noise using a non-i.i.d. mixture of Gaussians (NMoGs) noise assumption, which finely accords with the noise characteristics possessed by a natural HSI and thus is capable of adapting various practical noise shapes. Then we integrate such noise modeling strategy into the low-rank matrix factorization (LRMF) model and propose an NMoG-LRMF model in the Bayesian framework. A variational Bayes algorithm is then designed to infer the posterior of the proposed model. As substantiated by our experiments implemented on synthetic and real noisy HSIs, the proposed method performs more robust beyond the state-of-the-arts.

**Index Terms**—Hyperspectral image (HSI) denoising, low-rank matrix factorization (LRMF), non independent and identically distributed (i.i.d.) noise modeling.

## I. INTRODUCTION

**H**YPERSPECTRAL image (HSI) is captured from sensors over various bands and contains abundant spatial and spectral knowledge across all these bands beyond the traditional gray-scale or RGB images. Due to its preservation of full-bands information under a real scene, it has been widely used in military and civilian aspects such as terrain detection, mineral exploration, pharmaceutical counterfeiting, vegetation and environmental monitoring [16]. In real cases, however, an HSI is always corrupted by noises due to equipment limitations, such as sensor sensitivity, photon effects and

calibration error. Besides, due to the limited radiance energy and generally narrow band width, the energy captured by each sensor might be low and thus the shot noise and thermal noise then always happen inevitably. These noise severely degrades the quality of the imagery and limits the performance of the subsequent processing, e.g., classification [39], unmixing [2], and target detection [41], on data. Therefore, it is a critical pre-processing step to reduce the HSI noise [20], [38] to a general HSI processing task.

The simplest denoising way is to utilize the traditional 2-D or 1-D denoising methods to reduce noise in the HSI pixel by pixel [17] or band by band [12], [14], [29], [40], [46]. However, this kind of processing method ignores the correlations between different spectral bands or adjacent pixels, and often results in a relatively low-quality result. Recently, more efficient methods have been proposed to handle this issue. The main idea is to elaborately encode the prior knowledge on the structure underlying a natural HSI, especially the characteristic across the spatial and spectral dimensionality. For example, Othman and Qian [37] made an initial attempt to this issue by designing a hybrid spatial-spectral derivative-domain wavelet shrinkage model based on the dissimilarity of the signal regularity existing along the space and spectrum of a natural HSI. And then Chen and Qian [6] proposed another approach to encoding both spatial and spectral knowledge by combining the bivariate wavelet thresholding with principal component analysis. To further enhance the denoising capability, Yuan *et al.* [47] employed a spectral-spatial adaptive total variation. Later, Chen *et al.* [10] proposed a spatially adaptive weighted prior by combining the smoothness and discontinuity preserving properties along the spectral domain. By further considering the spatial and spectral dependencies, Zhong and Wang [52] proposed a multiple-spectral-band CRF (MSB-CRF) model in a unified probabilistic framework.

Besides, by explicitly treating HSI data as a tensor, a series of methods that expanding wavelet-based method from 2-D to 3-D has been proposed [45]. For example, Dabov *et al.* [11] designed VBM3D method by applying the concepts of grouping and collaborative filtering to video denoising. Then, Letexier and Bourennane [27] proposed a generalized multidimensional Wiener filter for denoising HSI. Similarly, Chen and Qian [7] extended Sendur and Selesnick's bivariate wavelet thresholding from 2-D image denoising to 3-D data cube denoising. For better denoising results, Chen and Zhu [8] proposed a new signal denoising algorithm by using neighboring wavelet coefficients, which considered both translation-invariant (TI) and

Manuscript received June 1, 2016; revised November 19, 2016; accepted February 15, 2017. This work was supported in part by the NSFC Projects under Grant 61373114, Grant 61661166011, Grant 11690011, Grant 61603292, Grant 11501440, and Grant 61503263, and in part by the 973 Program of China under Grant 3202013CB329404. This paper was recommended by Associate Editor L. Shao. (*Corresponding author: Deyu Meng.*)

The authors are with the School of Mathematics and Statistics and the Ministry of Education Key Laboratory of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dymeng@mail.xjtu.edu.cn).

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

Digital Object Identifier 10.1109/TCYB.2017.2677944

non-TI versions. Later, as an extension of BM3D method, Maggioni *et al.* [31] presented BM4D model. Meanwhile, another type of method that based on tensor decomposition has appeared. Karami *et al.* [25] developed a genetic kernel Tucker decomposition algorithm to exploit both spectral and the spatial information in HSIs. To address the uniqueness of multiple ranks of Tucker decomposition, Liu *et al.* [30] proposed PARAFAC method that make the number of estimated rank reduced to one. Later, Peng *et al.* [38] proposed a decomposable nonlocal tensor dictionary learning (TDL) model, which fully considered the nonlocal similarity over space and the global correlation across spectrum. Among these methods, BM4D and TDL achieved the state-of-the-art performance in more general noisy MSI cases.

Most of current HSI denoising methods have mainly considered the HSI prior spectral and spatial knowledge into their methods, while only use  $L_2$  loss term to rectify the deviation between the reconstruction and the original HSI. In the viewpoint of statistical theory, the utilization of such loss term implies that the HSI data noise follows an independent and identically distributed (i.i.d.) Gaussian distribution. However, in real scenarios, HSI noises generally have more complicated statistical structures. This means that such easy loss term is too subjective to reflect the real cases and inclines to degenerate the performance of the methods in more complex while more realistic non-i.i.d. noise case.

The idea of considering more complex HSI noise beyond only Gaussian in HSI denoising has been attracting attention very recently in the framework of low-rank matrix analysis. Since adjacent HSI bands usually exhibit strong correlations, by reshaping each band as a vector and stacking all the vectors into a matrix, the reshaped HSI matrix can be assumed to be with low rank. Various low-rank matrix models have been presented in different noise scenes in recent decades. Along this line, the classical low-rank matrix factorization (LRMF) model is presented [35], [36] for Gaussian noise, and its global optimal solution can be directly obtained by using singular value decomposition (SVD) [18]. To add more robustness, the  $L_1$ -norm LRMF [13], [15], [24], [26] is generally used and many algorithms have been designed to solve this model, such as  $L_1$  Wiberg [15], CWM [34], and RegL1ALM [51]. This  $L_1$ -norm LRMF model actually assumes an i.i.d. Laplacian noises embedded in data. To handle more complex noise cases, [33] modeled the noise as more adaptable i.i.d. mixture of Gaussians (MoGs) distributions to represent noise. Such noise modeling idea was further extended to the Bayesian framework in [9], to RPCA in [50] and to the tensor factorization in [44]. Very recently, Cao *et al.* [4], [5] modeled the noise as a more general i.i.d. mixture of exponential power (MoEP) distribution, which achieves competitive performance in HSI denoising under real noise scenarios. Besides, some other attempts have also been proposed to model the noise as a combination of sparse and Gaussian noise [1], and Zhang *et al.* [49] also utilized this idea for HSI denoising task. Later, He *et al.* [20] further enhanced the capability of the method by adding a TV regularization to low-rank reconstruction. Besides, based on the mixture noise assumption, He *et al.* [19] proposed a noise-adjusted low-rank methods

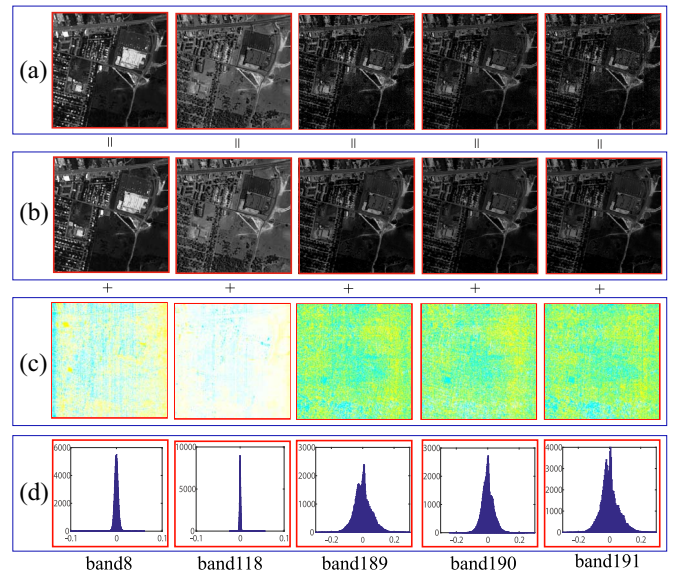


Fig. 1. (a) Different bands of images in the original HSI. (b) Restored HSIs obtained by our proposed method. (c) Extracted noise by the proposed method. (d) Histograms of the noise in all bands.

and Wang *et al.* [42] proposed a GLRR denoising method for the reconstruction of the corrupted HSIs. All these approaches also achieve good performance on HSI denoising in mixed noise cases.

From Gaussian noise assumption to MoEP noise assumption, such advancements make the model more adaptive to various HSI noises encountered in practice. However, all of the aforementioned LRMF models just simply assume that the HSI noise is i.i.d., which is more or less deviated from the practical scenarios, where the noises in an HSI is generally with non-i.i.d. configurations. In this sense, there is still a large room to further improve the capability of current HSI denoising techniques especially under real complicated scenarios.

To make this point clearer, let us try to analyze some evident noise characteristics possessed by a practical HSI. Fig. 1 presents an HSI for auxiliary understanding. First, an image in a band is generally collected under the similar sensor setting, and thus the i.i.d. distribution is a rational assumption for the noise over the elements in the band. This can be evidently observed from the band-noise of the HSI image shown in Fig. 1. Second, due to the difference of the sensor sensitivity for collecting images located in various HSI bands, the noise of different-band-images always depicts evident distinctions [19]. From Fig. 1, it is easy to see that images located in some bands are very clean, while are extremely noisy in some others [16], [48]. Third, albeit different, the noise distributions along different bands have certain correlation. For example, along adjacent bands, the noise tends to be more or less similar since images on neighboring spectrums are generally collected under small deviation of sensor settings and wavelength [52]. From Fig. 1, it is easy to see the noise similarity for images located in 189-191 bands. Accordingly, the real HSI noise distribution is generally non-i.i.d. and has a more complicated configurations than the i.i.d. noise assumption of the current HSI denoising techniques. Such deviation inclines to make their performance degenerate under

more practical cases, which will be clearly observed in our experiments.

To address this issue, this paper proposes a new noise modeling framework, by carefully designing noise distribution structure to make it possibly faithfully deliver the real HSI noise configurations. Specifically, we model the noise of each HSI band with different MoG distributions. Besides, MoG parameters across different bands are encoded under a similar top-level prior distribution, representing the correlation between noise distributions across different bands. In this way, the non-i.i.d. noise structure under a practical HSI image can be more properly encoded and an appropriate HSI denoising effect is thus to be expected.

In this paper, we embed such noise modeling framework into the LRMF realization, which easily assumes a low-rank structure (both spatial and spectral) of the to-be-recovered HSI. Our experimental results show that, even under this simple setting of HSI structure, such noise modeling idea can help evidently enhance the HSI denoising ability beyond previous techniques.

The rest of this paper is organized as follows. The proposed model and the corresponding variational inference algorithm are presented in Section II. Experimental results are shown in Section III. Finally, conclusions are drawn in Section IV. Throughout this paper, we denote scalars, vectors, matrices, and tensors as nonbold letters, bold lower case letters, bold upper case letters, and decorated letters, respectively.

## II. NON-I.I.D. MOG METHOD FOR HSI DENOISING

In this section, we first introduce our non-i.i.d. noise encoding strategy and then propose a non-i.i.d. MoG LRMF (NMoG-LRMF) model by using this strategy in the LRMF model for HSI denoising. Finally, the corresponding variational inference algorithm for the proposed model is designed.

### A. Non-i.i.d. Noise Encoding

Let us represent a given HSI as a matrix  $\mathbf{Y} \in \mathbb{R}^{N \times B}$ , where  $N$  and  $B$  mean the spatial and spectral dimensions of the HSI, respectively. By assuming that noise is additive, we have

$$\mathbf{Y} = \mathbf{L} + \mathbf{E} \quad (1)$$

where  $\mathbf{L} \in \mathbb{R}^{N \times B}$  denotes the clean HSI and  $\mathbf{E} \in \mathbb{R}^{N \times B}$  denotes the embedded noise.

Owing to the three intrinsic properties possessed by HSI noise.

- 1) Noise of an image located in the same band are i.i.d.
- 2) Noise of images located in different bands are nonidentical.
- 3) Noise distributions in different bands are correlated, we can present a proper model for noises in an HSI.

Via the above properties 1 and 2 of HSI noise structure, we model noise located in each band as an independent MoG distribution while assume that the parameters for MoG distributions in different bands are different. The MoG is utilized due to its universal approximation capability for any continuous densities [32], which has been extensively used and studied in [33] and [50].

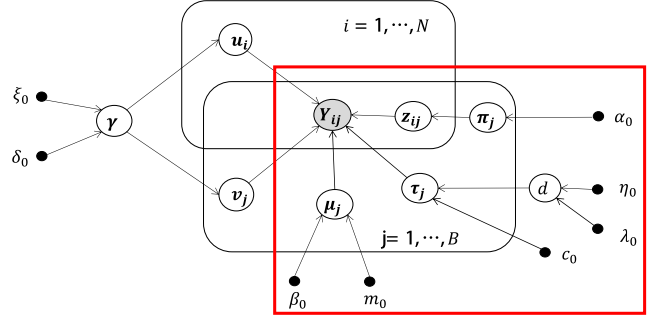


Fig. 2. Graphical model of NMoG-LRMF.  $Y_{ij}$  denotes the  $i$ th HSI element in its  $j$ th band.  $u_i$  and  $v_j$  are columns of low-rank matrix  $U$  and  $V$ , respectively, generated from a Gaussian distribution with precise  $\gamma$ , with Gamma prior distribution with hyper-parameters  $\xi_0$  and  $\delta_0$ . The portion in the red box corresponds to the noise encoding part.  $\mu_j$ ,  $\tau_j$  are the  $k$ -dimensional vectors representing the mean and variance of all MoG components in the  $j$ th band, with hyper-parameters  $\beta_0$ ,  $m_0$ ,  $d$ ,  $c_0$ , where  $d$  is with hyper-parameters  $\eta_0$ ,  $\lambda_0$ .  $z_{ij}$  is the hidden variable generated from Multinomial distribution with parameter  $\pi_j$ , with hyper-parameters  $\alpha_0$ .

Express  $e_{ij}$  as the element located in  $i$ th row and  $j$ th column of the noise matrix  $\mathbf{E}$ , and as aforementioned, the noise distribution located in the  $j$ th band can be modeled as

$$p(e_{ij}) = \sum_{k=1}^K \pi_{jk} N(e_{ij} | \mu_{jk}, \tau_{jk}^{-1}) \quad (2)$$

where  $\pi_{jk}$  is the mixing proportion with  $\pi_{jk} \geq 0$  and  $\sum_{k=1}^K \pi_{jk} = 1$ ,  $K$  is the Gaussian component number,  $\mu_{jk}$  and  $\tau_{jk}$  are mean and precision of the  $k$ th Gaussian component in the  $j$ th band, respectively. Note that MoG parameters  $\pi_{jk}$ ,  $\mu_{jk}$ , and  $\tau_{jk}$  are different, implying different MoG distributions across different bands.

We then consider the noise property 3, and can further provide the hypothesis that MoG parameters  $\mu_{js}$  and  $\tau_{js}$  of all bands are generated from a two-level prior distribution

$$\begin{aligned} \mu_{jk}, \tau_{jk} &\sim N(\mu_{jk} | m_0, (\beta_0 \tau_{jk})^{-1}) \text{Gam}(\tau_{jk} | c_0, d) \\ d &\sim \text{Gam}(d | \eta_0, \lambda_0) \end{aligned} \quad (3)$$

where  $\text{Gam}(\cdot)$  represents the Gamma distribution. In this way, the correlation between noise distributions among different bands is then rationally encoded. Through introducing a latent variable  $z_{ijk}$ , we can equivalently rewrite (2) as

$$\begin{aligned} e_{ij} &\sim \prod_{k=1}^K N(e_{ij} | \mu_{jk}, \tau_{jk}^{-1})^{z_{ijk}} \\ z_{ij} &\sim \text{Multinomial}(z_{ij} | \pi_j) \\ \pi_j &\sim \text{Dir}(\pi_j | \alpha_0) \end{aligned} \quad (4)$$

where  $\text{Multinomial}(\cdot)$  and  $\text{Dir}(\cdot)$  represent the multinomial and Dirichlet distributions, respectively. Then, (3) and (4) together encode the noise structure embedded in an HSI. Fig. 2 shows the graphical model for noise encoding within the red box. All involved parameters can be inferred from data, as introduced in Section II-C.



### B. NMoG-LRMF Model

Readily employing the low-rank structure of a clean HSI, we can encode the prior structure of the latent HSI matrix  $\mathbf{L}$ . Specifically, let us consider the following LRMF model:

$$\mathbf{Y} = \mathbf{UV}^T + \mathbf{E} \quad (5)$$

where  $\mathbf{L} = \mathbf{UV}^T$  implies the low-rank structure underlying the clean HSI.

For most deterministic LRMF model, the rank  $r$  of matrix  $\mathbf{L}$  is fixed. By modeling the problem into certain generative model [1], [50], the rank can also be adaptively learned from data. Specifically, suppose the columns of  $\mathbf{U}$  and  $\mathbf{V}$  are generated from Gaussian distribution. For  $l = 1, 2, \dots, R$ , where  $R$  is a preset larger value beyond the true rank  $r$ , then

$$\mathbf{u}_{\cdot l} \sim N(\mathbf{u}_{\cdot l} | 0, \gamma_l^{-1} \mathbf{I}_N), \quad \mathbf{v}_{\cdot l} \sim N(\mathbf{v}_{\cdot l} | 0, \gamma_l^{-1} \mathbf{I}_B) \quad (6)$$

where  $\mathbf{u}_{\cdot l}$  and  $\mathbf{v}_{\cdot l}$  are the  $l$ th columns of  $\mathbf{U}$  and  $\mathbf{V}$ , respectively.  $\mathbf{I}_N(\mathbf{I}_B)$  denotes the  $N \times N$  ( $B \times B$ ) identity matrix.  $\gamma_l$  is the precision of  $\mathbf{u}_{\cdot l}$  and  $\mathbf{v}_{\cdot l}$  with prior as follows:

$$\gamma_l \sim \text{Gam}(\gamma_l | \xi_0, \delta_0). \quad (7)$$

Note that each column pair  $\mathbf{u}_{\cdot l}$  and  $\mathbf{v}_{\cdot l}$  of  $\mathbf{U}$ ,  $\mathbf{V}$  has the same sparsity profile characterized by the common precision variable  $\gamma_l$ . It has been validated that such a modeling could lead to large precision values of some  $\gamma_l$ s, and hence is capable of automatically conducting low-rank estimate of  $\mathbf{L}$  [1].

Combining (2)–(7) together, we can construct the full NMoG-LRMF Bayesian model. And the graphical model representation of this model is shown in Fig. 2. The goal turns to infer the posterior of all involved variables

$$p(\mathbf{U}, \mathbf{V}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\pi}, \boldsymbol{\gamma}, d | \mathbf{Y}) \quad (8)$$

where  $\mathbf{Z} = \{\mathbf{z}_{ij}\}_{N \times B}$ ,  $\boldsymbol{\mu} = \{\mu_{jk}\}_{B \times K}$ ,  $\boldsymbol{\tau} = \{\tau_{jk}\}_{B \times K}$ ,  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_B)$ , and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_R)$ .

### C. Variational Inference

We use variational Bayes (VB) method [3] for posterior inference. Specifically, VB aims to use a variational distribution  $q(\boldsymbol{\theta})$  to approximate the true posterior  $p(\boldsymbol{\theta} | \mathbf{D})$ , where  $\boldsymbol{\theta}$  denotes the set of parameters and  $\mathbf{D}$  denotes the observed data. To achieve this goal, we need to solve the following optimization problem:

$$\min_{q \in \mathbb{C}} KL(q(\boldsymbol{\theta}) || p(\boldsymbol{\theta} | \mathbf{D})) = - \int q(\boldsymbol{\theta}) \ln \left\{ \frac{p(\boldsymbol{\theta} | \mathbf{D})}{q(\boldsymbol{\theta})} \right\} d\boldsymbol{\theta} \quad (9)$$

where  $KL(q || p)$  represents the KL divergence between two distributions  $q$  and  $p$ , and  $\mathbb{C}$  denotes the constraint set of probability densities to make the minimization tractable. Assuming  $\mathbb{C}$  is the distribution family which can be factorized with respect to some disjoint groups:  $q(\boldsymbol{\theta}) = \prod_i q_i(\boldsymbol{\theta}_i)$ , the closed-form optimal solution  $q_i^*(\boldsymbol{\theta}_i)$  can be obtained by

$$q_i^*(\boldsymbol{\theta}_i) = \frac{\exp\{\langle \ln p(\boldsymbol{\theta}, \mathbf{D}) \rangle_{\boldsymbol{\theta} \setminus \boldsymbol{\theta}_i}\}}{\int \exp\{\langle \ln p(\boldsymbol{\theta}, \mathbf{D}) \rangle_{\boldsymbol{\theta} \setminus \boldsymbol{\theta}_i}\}} \quad (10)$$

where  $\langle \cdot \rangle$  denotes the expectation and  $\boldsymbol{\theta} \setminus \boldsymbol{\theta}_i$  denotes the set of  $\boldsymbol{\theta}$  with  $\boldsymbol{\theta}_i$  removed.

Then we can analytically infer all the factorized distributions involved in (8). Suppose that the approximation of posterior distribution (8) possesses a factorized form as follows:

$$q(\mathbf{U}, \mathbf{V}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\pi}, \boldsymbol{\gamma}, d) = \prod_i q(\mathbf{u}_{\cdot i}) \prod_j q(\mathbf{v}_{\cdot j}) \prod_{ij} q(\mathbf{z}_{ij}) \\ \times \prod_j q(\boldsymbol{\mu}_j, \boldsymbol{\tau}_j) q(\boldsymbol{\pi}_j) \prod_l q(\gamma_l) q(d) \quad (11)$$

where  $\mathbf{u}_{\cdot i}$  ( $\mathbf{v}_{\cdot j}$ ) are the  $i$ th ( $j$ th) row of matrix  $\mathbf{U}$  ( $\mathbf{V}$ ), respectively. According to (10), we can get the closed-form inference equation for each component of (11).<sup>1</sup>

1) *Estimation of Noise Component:* We first list the updating equation for the noise components involved in (11). The posterior distribution of mean and precision for noise in each band is updated by the following equation:

$$q^*(\boldsymbol{\mu}_j, \boldsymbol{\tau}_j) = \prod_k N\left(\boldsymbol{\mu}_{jk} | m_{jk}, \frac{1}{\beta_{jk} \tau_{jk}}\right) \text{Gam}(\tau_{jk} | c_{jk}, d_{jk}) \quad (12)$$

where the parameters in the above equation can be calculated from data in the following way:

$$m_{jk} = \frac{1}{\beta_{jk}} \left\{ +\beta_0 m_0 \sum_i \langle z_{ijk} \rangle (Y_{ij} - \langle \mu_{jk} \rangle) \right\} \quad (13)$$

$$\beta_{jk} = \beta_0 + \sum_i \langle z_{ijk} \rangle, \quad c_{jk} = c_0 + \frac{1}{2} \sum_i \langle z_{ijk} \rangle \quad (14)$$

$$d_{jk} = \langle d \rangle + \frac{1}{2} \left\{ \sum_i \langle z_{ijk} \rangle \left\langle \left( Y_{ij} - \mathbf{u}_{\cdot i} \mathbf{v}_{\cdot j}^T \right)^2 \right\rangle + \beta_0 m_0^2 \right. \\ \left. - \frac{1}{\beta_{jk}} \left[ \sum_i \langle z_{ijk} \rangle \left( Y_{ij} - \mathbf{u}_{\cdot i} \mathbf{v}_{\cdot j}^T \right) + \beta_0^2 m_0^2 \right]^2 \right\}. \quad (15)$$

The update equation of the latent variable  $\mathbf{Z}$  is

$$q^*(\mathbf{z}_{ij}) = \prod_k q_{ijk}^{z_{ijk}} \quad (16)$$

where the involved parameters are calculated by

$$q_{ijk} = \rho_{ijk} / \sum_k \rho_{ijk} \\ \ln \rho_{ijk} = \langle \ln \pi_{jk} \rangle - \ln \sqrt{2\pi} + \langle \ln \tau_{jk} \rangle / 2 \\ - \left\langle \tau_{jk} \left( Y_{ij} - \mu_{jk} - \mathbf{u}_{\cdot i} \mathbf{v}_{\cdot j}^T \right)^2 \right\rangle / 2. \quad (17)$$

Similarly, the update equation for the mixing proportion  $\boldsymbol{\pi}_j$  over the  $j$ th band can be written as

$$q^*(\boldsymbol{\pi}_j) = \prod_k \pi_{jk}^{\alpha_{jk}-1} \quad (18)$$

where  $\alpha_{jk} = \alpha_0 + \sum_i \langle z_{ijk} \rangle$ .

The update equation on the hyper-parameter  $d$  is

$$q^*(d) = \text{Gam}(d | \eta, \lambda) \quad (19)$$

where  $\eta = \eta_0 + c_0 KB$  and  $\lambda = \lambda_0 + \sum_{j,k} \langle \tau_{jk} \rangle$ .

<sup>1</sup>Inference details to obtain these equations can be referred to in <http://dymeng.gr.xjtu.edu.cn>.

2) *Estimation of Low-Rank Component*: Completing the component update for noise, we then estimate the posterior of low-rank component  $\mathbf{u}_i$  ( $i = 1, \dots, N$ ) and  $\mathbf{v}_j$  ( $j = 1, \dots, B$ ) as

$$q^*(\mathbf{u}_i) = N(\mathbf{u}_i | \boldsymbol{\mu}_{\mathbf{u}_i}, \boldsymbol{\Sigma}_{\mathbf{u}_i}) \quad (20)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{u}_i} &= \left\{ \sum_{j,k} \langle z_{ijk} | \tau_{jk} \rangle (Y_{ij} - \langle \mu_{jk} \rangle) \langle \mathbf{v}_j \rangle \right\} \boldsymbol{\Sigma}_{\mathbf{u}_i} \\ \boldsymbol{\Sigma}_{\mathbf{u}_i} &= \left\{ \sum_{j,k} \langle z_{ijk} | \tau_{jk} \rangle \langle \mathbf{v}_j^T \mathbf{v}_j \rangle + \langle \Gamma \rangle \right\}^{-1} \\ q^*(\mathbf{v}_j) &= N(\mathbf{v}_j | \boldsymbol{\mu}_{\mathbf{v}_j}, \boldsymbol{\Sigma}_{\mathbf{v}_j}) \end{aligned} \quad (21)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{v}_j} &= \left\{ \sum_{i,k} \langle z_{ijk} | \tau_{jk} \rangle (Y_{ij} - \langle \mu_{jk} \rangle) \langle \mathbf{u}_i \rangle \right\} \boldsymbol{\Sigma}_{\mathbf{v}_j} \\ \boldsymbol{\Sigma}_{\mathbf{v}_j} &= \left\{ \sum_{i,k} \langle z_{ijk} | \tau_{jk} \rangle \langle \mathbf{u}_i^T \mathbf{u}_i \rangle + \langle \Gamma \rangle \right\}^{-1} \end{aligned} \quad (22)$$

where  $\Gamma = \text{diag}(\langle \gamma \rangle)$ . For  $\gamma_l$  which controls the rank of  $\mathbf{U}$  and  $\mathbf{V}$ , we have

$$q^*(\gamma_l) = \text{Gam}(\gamma_l | \xi_l, \delta_l) \quad (23)$$

where

$$\begin{aligned} \xi_l &= \xi_0 + (m + n)/2 \\ \delta_l &= \delta_0 + \sum_i \langle u_{il}^2 \rangle / 2 + \sum_j \langle v_{jl}^2 \rangle / 2 \end{aligned}$$

$m$  and  $n$  represent the image size among the spatial dimension. As discussed by Babacan *et al.* [1], some  $\gamma_l$ s tend to be very large during the inference process and the corresponding rows will be removed from  $\mathbf{U}$  and  $\mathbf{V}$ . The low-rank purpose can thus be rationally conducted. In all our experiments, we just automatically infer the rank of the reconstructed matrix through throwing away those comparatively very large  $\gamma_l$  as previous literatures did [50].

The proposed variational inference method for the NMoG-LRMF model can then be summarized in Algorithm 1.

3) *Setting of Hyper-Parameters*: We set all the hyper-parameters involved in our model in a noninformative manner to make them possibly less affect the inference of posterior distributions [3]. Throughout our experiments, we set  $m_0$  as 0, and  $\alpha_0, \beta_0, c_0, d_0, \eta_0, \lambda_0, \xi_0, \delta_0$  as a small value  $10^{-3}$ . Our method performs stably well under such easy settings.

### III. EXPERIMENTAL RESULTS

In this section, to evaluate the performance of the proposed NMoG-LRMF method, we conducted a series of experiments on both synthetic and real HSI data. Compared methods include LRMR [49] and LRTV [20], considering deterministic Gaussian noise and sparse noise. Meanwhile, five representative low-rank matrix analysis methods considering

#### Algorithm 1 NMoG-LRMF Algorithm

**Input**: the original HSI matrix  $\mathbf{Y} \in \mathbb{R}^{N \times B}$ , Gaussian component number  $K$ , and maximum iteration number.

**Output**:  $\mathbf{U}_{\text{opt}} = \mathbf{U}_t, \mathbf{V}_{\text{opt}} = \mathbf{V}_t$ .

**Initialization**: Parameters  $(m_0, \beta_0, c_0, d_0, \eta_0, \lambda_0)$  in noise prior. Low-rank components  $\mathbf{U}_0, \mathbf{V}_0$  and  $\alpha_0$  parameters in model prior  $(\xi_0, \delta_0)$ ;  $t = 1$ .

- 1: **while** not covered **do**
- 2:   Update approximate posterior of noise component  $\mathcal{Z}^t, \boldsymbol{\pi}^t$  by Eq. (16)-(18).
- 3:   Update approximate posterior of noise component  $\boldsymbol{\mu}^t, \boldsymbol{\tau}^t$  by Eq. (12)-(15).
- 4:   Update approximate posterior of noise component  $d^t$  by Eq. (19).
- 5:   Update approximate posterior of low-rank component  $\mathbf{U}^t, \mathbf{V}^t$  by Eq. (20)-(22).
- 6:   Update approximate posterior of parameters in noise component  $\boldsymbol{\gamma}^t$  by Eq. (23).
- 7:    $t \leftarrow t + 1$ .
- 8: **end while**

different kinds of i.i.d. noise distributions were also considered for comparison, including PMoEP [4] (assuming i.i.d. MoEP noise), MoG-RPCA [50] (assuming i.i.d. MoG noise), RegL1ALM [51], CWM [34] (assuming i.i.d. Laplace noise), and SVD [18] (assuming i.i.d. Gaussian noise). Besides, the performance of TDL [38] and BM4D [31] are also compared, and both methods represent the state-of-the-art methods for HSI denoising by considering HSI priors. All experiments were implemented in MATLAB R2014b on a PC with 3.4-GHz CPU and 32-GB RAM.

#### A. Simulated HSI Denoising Experiments

In this experiment, we focus on the performance of NMoG-LRMF in HSI denoising with synthetic noise. Two HSIs were employed: Washington DC Mall<sup>2</sup> with size of  $1208 \times 307 \times 191$  and RemoteImage<sup>3</sup> provided by Liu *et al.* [28] with size of  $205 \times 246 \times 96$ . After cropping the main part of HSI and deleting some evident visual contaminative spectral channels, Washington DC Mall and RemoteImage are resized to  $200 \times 200 \times 160$  and  $200 \times 200 \times 89$ , respectively. The gray value of HSIs are normalized into  $[0, 1]$ .

Real-world HSIs are usually contaminated by several different types of noise, including the most common Gaussian noise, impulse noise, dead pixels or lines, and stripes [49]. In order to simulate these real HSI noise scenarios, we added six kinds of noises to the original HSI data.

- 1) *i.i.d. Gaussian Noise*: Entries in all bands were corrupted by zero-mean i.i.d. Gaussian noise  $N(0, \sigma^2)$  with  $\sigma = 0.05$ .
- 2) *Non-i.i.d. Gaussian Noise*: Entries in all bands were corrupted by zero-mean Gaussian noise with different intensity. The signal noise ratio (SNR) value of each band is generated from uniform distribution with value in the range of  $[10, 20]$ dB.
- 3) *Gaussian + Stripe Noise*: All bands were corrupted by Gaussian noise as case 2. Besides, 40 bands in DCmall

<sup>2</sup><http://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

<sup>3</sup>[http://peterwonka.net/Publications/code/LRTC\\_Package\\_Ji.zip](http://peterwonka.net/Publications/code/LRTC_Package_Ji.zip)

TABLE I  
PERFORMANCE COMPARISON OF ALL COMPETING METHODS ON DCMALL DATA

	Noisy HSI	SVD	RegL1ALM	CWM	MoG-RPCA	PMoEP	LRMR	LRTV	TDL	BM4D	NMoG
i.i.d. Gaussian Noise											
MPSNR	26.02	39.28	37.97	38.08	39.48	39.28	39.31	38.75	<b>40.93</b>	37.47	39.49
MSSIM	0.763	0.986	0.980	0.981	0.987	0.986	0.987	0.985	<b>0.991</b>	0.978	0.987
time	-	<b>0.501</b>	232.5	909.2	321.7	116.3	364.9	118.4	183.8	1367.0	334.4
Non i.i.d. Gaussian Noise											
MPSNR	24.86	36.98	36.88	36.67	37.82	37.37	38.05	37.41	30.19	36.74	<b>38.87</b>
MSSIM	0.733	0.976	0.974	0.973	0.981	0.977	0.982	0.980	0.899	0.978	<b>0.985</b>
time	-	<b>0.419</b>	207.1	797.5	274.5	997.4	362.5	172.3	746.4	1527.6	287.8
Gaussian + Stripe Noise											
MPSNR	24.35	36.58	36.62	36.46	37.55	37.18	37.46	37.16	28.02	35.17	<b>38.52</b>
MSSIM	0.716	0.974	0.972	0.971	0.980	0.976	0.980	0.978	0.844	0.965	<b>0.983</b>
time	-	<b>0.515</b>	238.5	782.6	307.6	1505.5	377.1	192.0	843.1	589.3	334.7
Gaussian + Deadline noise											
MPSNR	24.16	34.79	36.67	35.72	36.66	36.04	37.15	36.99	27.82	32.92	<b>38.31</b>
MSSIM	0.715	0.966	0.973	0.970	0.978	0.973	0.978	0.978	0.851	0.943	<b>0.983</b>
time	-	<b>0.536</b>	245.7	900.0	341.8	896.6	385.8	277.2	727.3	335.6	346.1
Gaussian + Impluse Noise											
MPSNR	24.19	35.73	39.52	37.37	38.91	36.58	39.94	39.95	27.92	34.20	<b>41.51</b>
MSSIM	0.724	0.971	0.987	0.981	0.986	0.976	0.989	0.989	0.851	0.955	<b>0.993</b>
time	-	<b>0.356</b>	197.8	671.8	260.4	773.9	343.8	229.9	554.8	376.3	273.0
Mixture Noise											
MPSNR	22.87	33.13	35.17	34.11	35.22	34.31	35.97	35.79	25.08	31.22	<b>37.02</b>
MSSIM	0.660	0.955	0.963	0.962	0.970	0.964	0.971	0.971	0.754	0.918	<b>0.977</b>
time	-	<b>0.323</b>	186.1	637.3	250.5	1263.7	395.1	256.1	588.0	493.3	258.3

TABLE II  
PERFORMANCE COMPARISON OF ALL COMPETING METHODS ON REMOTEIMAGE DATA

	Noisy HSI	SVD	RegL1ALM	CWM	MoG-RPCA	PMoEP	LRMR	LRTV	TDL	BM4D	NMoG
i.i.d. Gaussian Noise											
MPSNR	26.02	36.53	35.43	35.55	36.65	36.53	37.32	35.73	<b>38.73</b>	36.99	36.68
MSSIM	0.695	0.967	0.955	0.957	0.969	0.967	0.968	0.959	<b>0.975</b>	0.963	0.969
time	-	<b>0.544</b>	139.6	408.0	302.7	88.0	144.5	136.8	122.7	334.4	322.5
Non i.i.d. Gaussian Noise											
MPSNR	27.04	36.87	36.03	36.12	36.98	36.88	<b>37.97</b>	36.27	37.38	37.40	37.25
MSSIM	0.735	0.971	0.962	0.963	0.972	0.971	0.973	0.965	0.968	0.966	<b>0.974</b>
time	-	<b>0.451</b>	120.2	361.2	270.4	147.7	149.2	186.9	231.6	229.7	275.4
Gaussian + Stripe Noise											
MPSNR	25.81	35.81	35.78	35.80	36.66	36.42	<b>37.37</b>	35.99	31.71	33.16	37.01
MSSIM	0.697	0.963	0.960	0.961	0.970	0.965	0.971	0.963	0.864	0.907	<b>0.973</b>
time	-	<b>0.436</b>	120.9	372.4	290.2	1538.7	151.2	155.7	483.3	260.1	298.6
Gaussian + Deadline noise											
MPSNR	24.75	32.31	35.77	34.64	36.82	35.11	36.42	36.18	30.09	31.29	<b>37.01</b>
MSSIM	0.681	0.934	0.961	0.954	0.971	0.959	0.964	0.949	0.849	0.879	<b>0.972</b>
time	-	<b>0.487</b>	121.7	352.1	284.3	231.4	155.7	100.8	444.1	237.5	290.9
Gaussian + Impluse Noise											
MPSNR	24.15	32.23	37.32	35.28	37.14	35.07	38.88	37.74	29.37	30.66	<b>41.38</b>
MSSIM	0.605	0.916	0.972	0.959	0.973	0.958	0.977	0.974	0.759	0.839	<b>0.981</b>
time	-	<b>0.526</b>	130.9	370.6	277.7	1142.6	138.6	159.4	534.4	166.2	295.2
Mixture Noise											
MPSNR	22.41	29.64	32.92	33.18	34.75	33.29	34.53	34.06	26.20	27.32	<b>35.89</b>
MSSIM	0.555	0.868	0.918	0.934	0.953	0.931	0.942	0.944	0.725	0.735	<b>0.965</b>
time	-	<b>0.352</b>	105.6	321.6	244.3	306.9	130.4	125.0	255.8	172.4	255.2

data (30 bands in RemoteImage data) were randomly chosen to add stripe noise. The number of stripes in each band is from 20 to 40.

- 4) *Gaussian + Deadline Noise*: Each band was contaminated by Gaussian noise as case 2. Forty bands in DCmall data (30 bands in RemoteImage data) were chosen randomly to add deadline noise. The number of deadline is from 5 to 15.
- 5) *Gaussian + Impulse Noise*: All bands were corrupted by Gaussian noise as case 2. Forty bands in DCmall (30 bands in RemoteImage data) were randomly chosen to added impulse noise with different intensity, and the percentage of impulse is from 50% to 70%.

- 6) *Mixture Noise*: Each band was randomly corrupted by at least one kind of noise mentioned in case 2–5.

Three criteria were utilized to measure performance.

- 1) *MPSNR* [23]: Mean of peak SNR (PSNR) over all bands between clean HSI and recovered HSI.
- 2) *MSSIM* [43]: Mean of structural similarity (SSIM) between clean HSI and recovered HSI over all bands.
- 3) *Time*: Time cost of each method used to complete the denoising process.

The parameters of competing methods are set as follows: the block size in LRMR is  $20 \times 20 \times B$  and the step size is 4. For LRTV,  $\epsilon_1 = \epsilon_2 = 10^{-8}$  and  $\lambda = 1/\sqrt{NB}$ . For TDL, we set the block size as  $6 \times 6 \times B$  and the step size is 2. For BM4D,

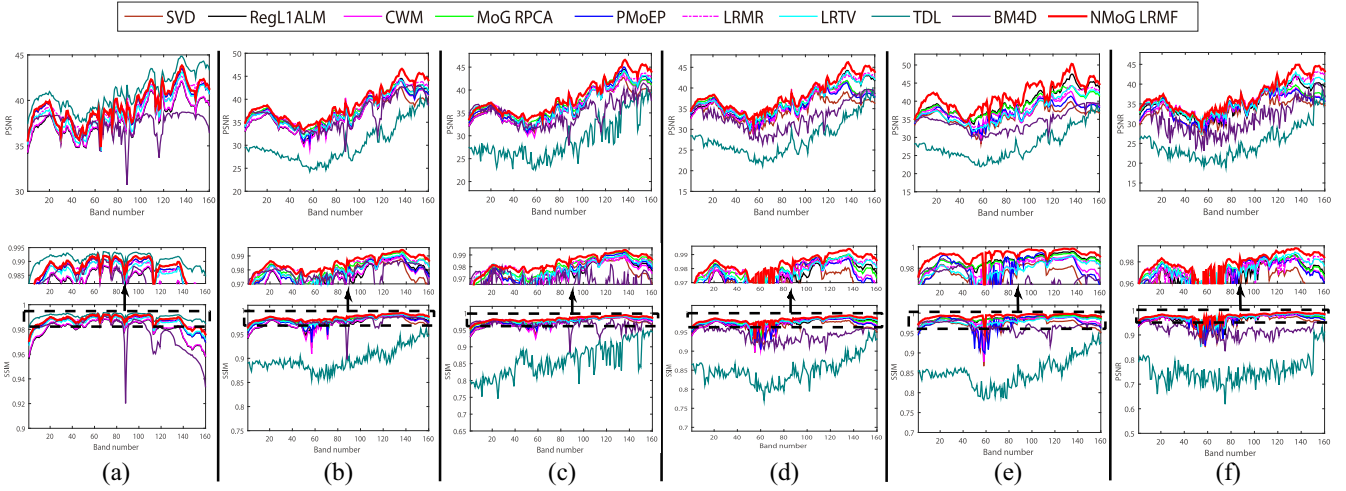


Fig. 3. Each column shows the average PSNR and SSIM measurements among 20 initializations of all methods under certain type of noise in DCmall data. (a) Gaussian noise. (b) Gaussian + stripe noise. (c) Gaussian + deadline noise. (d) Gaussian + impulse noise. (e) Mixture noise. The demarcated area of the subfigure indicates the curve locality on a larger scale.

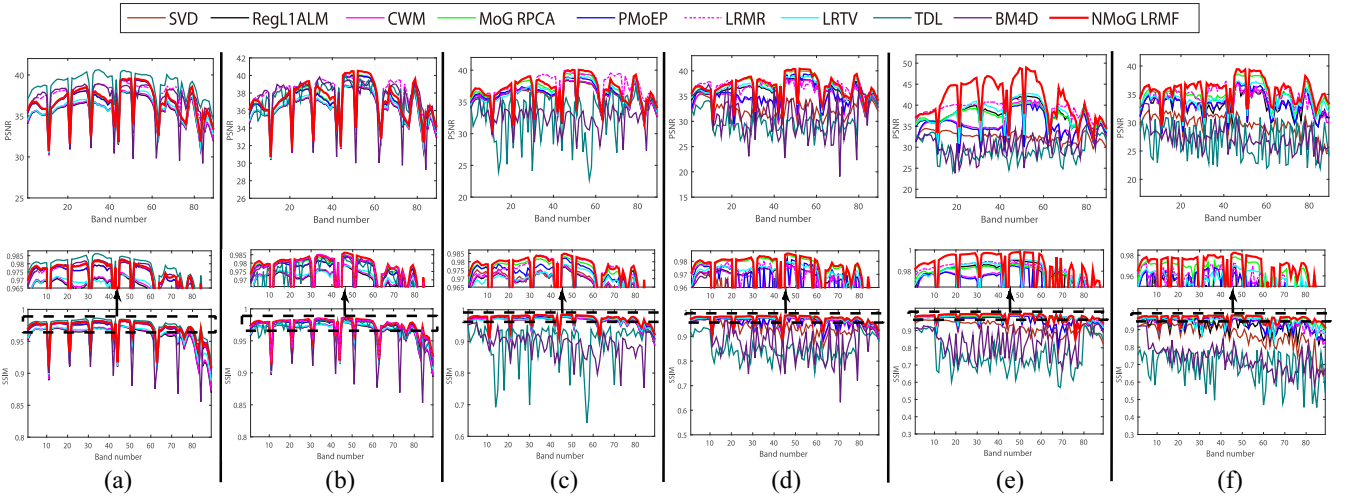


Fig. 4. Each column shows the average PSNR and SSIM measurements among 20 initializations of all methods under certain type of noise in RemoteImage data. (a) Gaussian noise. (b) Gaussian + Stripe noise. (c) Gaussian + deadline noise. (d) Gaussian + impulse noise. (e) Mixture noise.

we set the block size as  $8 \times 8 \times 64$  and the step size is 4. For NMoG-LRMF method, The component number  $K$  in each band was fixed as 1 in cases 1 and 2, and 3 in cases 3–6. The rank of all low-rank based methods is set as 5 in DCmall data experiment and 4 in RemoteImage data experiment. All parameters involved in the competing methods were carefully tuned or specified as suggested by the related literatures to guarantee their possibly good performance. All competing methods run with 20 random initializations in each noise case, and the average result is reported.

The results of all competing methods in DCmall and RemoteImage HSI data are shown in Tables I and II, respectively. The superiority of the proposed method can be easily observed, except in the i.i.d. Gaussian noise case, which complies with the basic noise assumption of conventional methods. In i.i.d. Gaussian case, instead of only using the simple low-rank prior in our method, multiple competing methods, like TDL, BM4D and LRMR, utilize more useful

HSI priors in their model, and thus tend to have relatively better performance. While on more complex but more practical complicated non-i.i.d. noise cases, the advantage of the proposed method is evident. This can be easily explained by the better noise fitting capability of the proposed method, i.e., it can more properly extract noises embedded in HSI, which then naturally leads to its better HSI recovery performance.

Furthermore, it also can be seen that the computational cost of the proposed method is with almost similar order of magnitude with other competing methods, except the known SVD, which we use the mature toolkit in MATLAB and can be implemented very efficiently. Considering its better capability in fitting much wider range of noises than current methods, it should be rational to say that the proposed method is efficient.

We further show the PSNR and SSIM measurements across all bands of the HSI under six types of noise settings in two experiments in Figs. 3 and 4, respectively. From the figures, it is easy to see that TDL obtains the best PSNR and



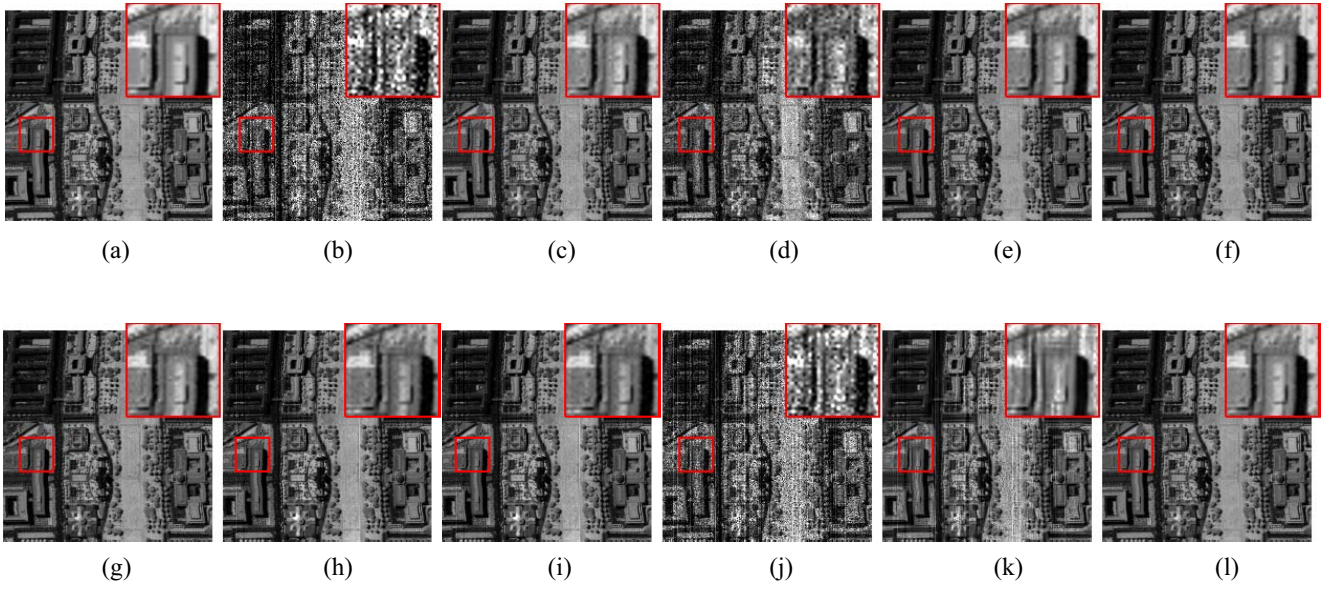


Fig. 5. Restoration results of band 75 under mixture noise in DCmall data. (a) Original HSI. (b) Noisy HSI. (c) SVD. (d) RegL1ALM. (e) CWM. (f) MoG-RPCA. (g) PMoEP. (h) LRMR. (i) LRTV. (j) TDL. (k) BM4D. (l) NMoG.

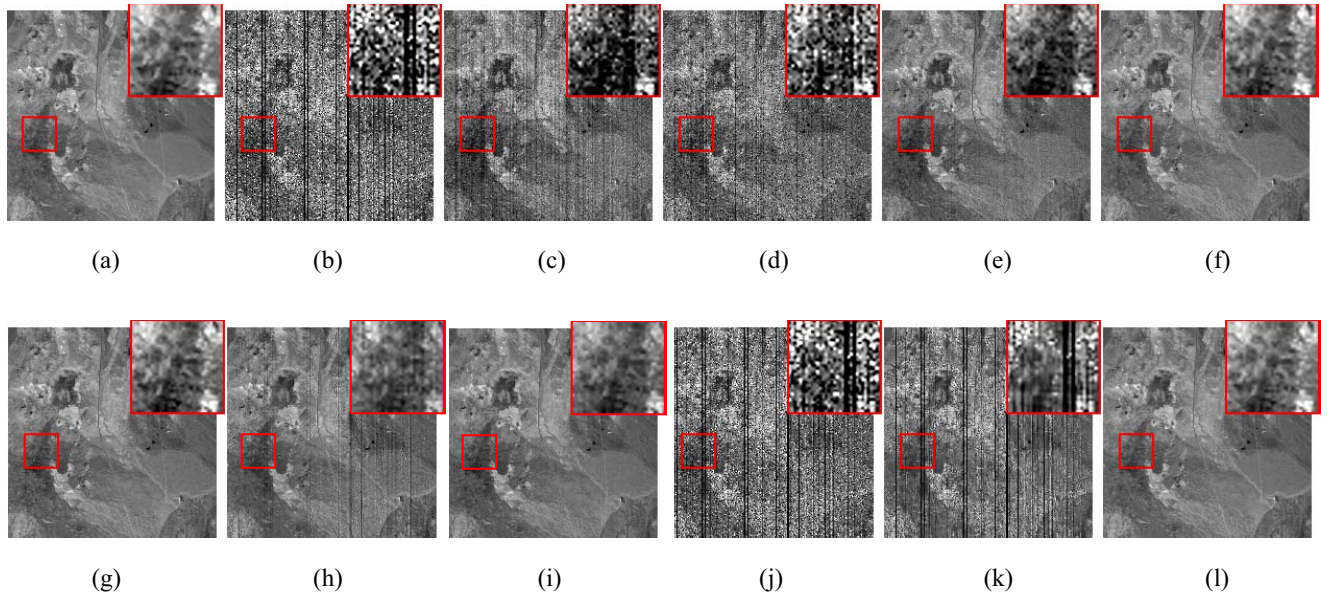


Fig. 6. Restoration results of band 86 under mixture noise in RemoteImage data. (a) Original HSI. (b) Noisy HSI. (c) SVD. (d) RegL1ALM. (e) CWM. (f) MoG-RPCA. (g) PMoEP. (h) LRMR. (i) LRTV. (j) TDL. (k) BM4D. (l) NMoG.

SSIM values across all bands in the i.i.d. Gaussian noise. In overwhelming majority of other more complex noise cases, NMoG-LRMF achieves the best PSNR and SSIM values across almost all bands. This verifies the robustness of the proposed method over entire HSI bands.

Figs. 5 and 6 give the restoration results of two typical bands in DCmall and RemoteImage HSIs, respectively. The original HSIs are corrupted by MoG, stripes, deadline, and impulse noise. It can be observed that the competing methods SVD, RegL1ALM, TDL, and BM4D can hardly remove this complex mixture noise from the images. We can also see that although CWM, MoG-RPCA, PMoEP, LRMR, and LRTV have a better denoising performance, the restored HSI still

relatively blur many details or maintain some noise as compared with the ground truth. Comparatively, our NMoG-LRMF method achieves better reconstruction effect both visually and quantitatively, and both finely removes the unexpected noises, and better recovers HSI details like edges and textures.

### B. Real HSI Denoising Experiments

In this section, we evaluate the performance of the proposed method on two real HSI datasets, Urban dataset<sup>4</sup> and

<sup>4</sup><http://www.tec.army.mil/hypercube>



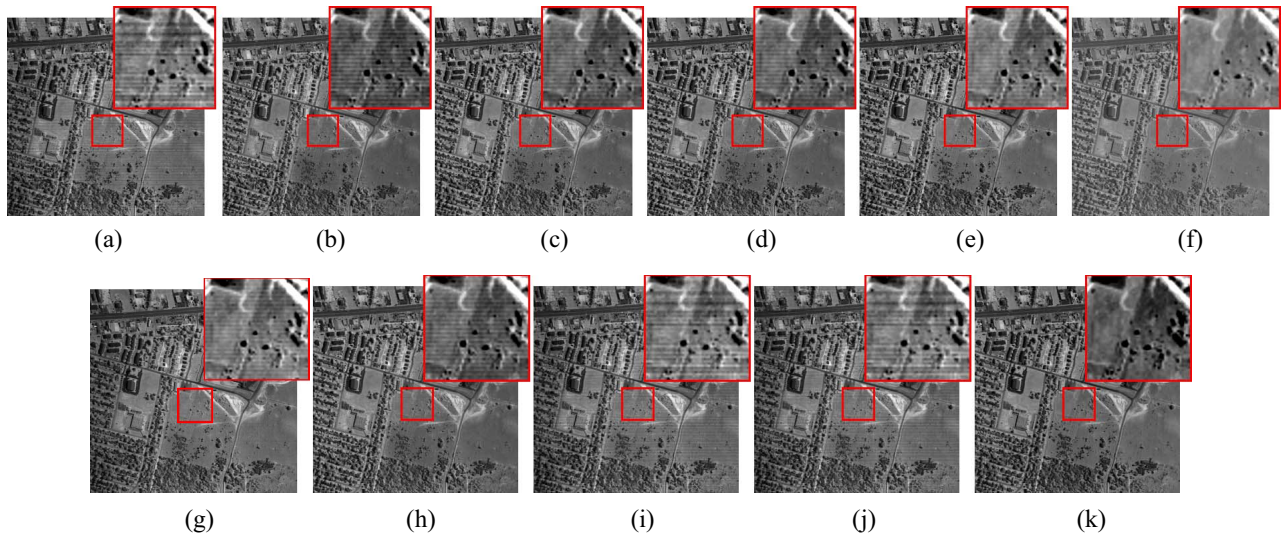


Fig. 7. Restoration results of all methods on band 103 in Urban HSI data. (a) Original HSI. (b) SVD. (c) RegL1ALM. (d) CWM. (e) MoG-RPCA. (f) PMoEP. (g) LRMR. (h) LRTV. (i) TDL. (j) BM4D. (k) NMoG.

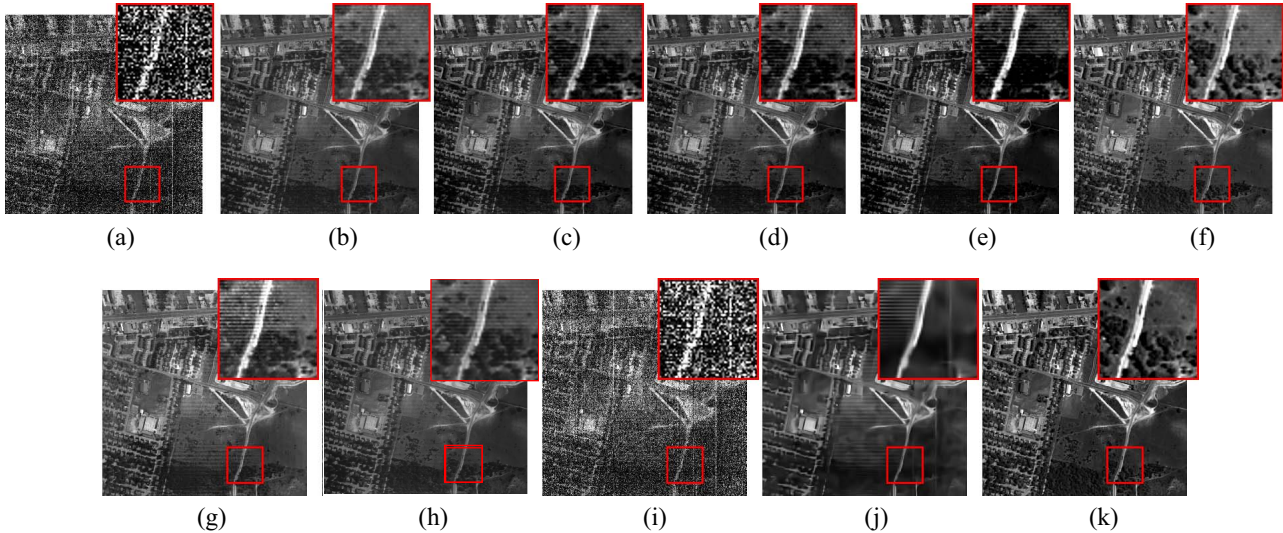


Fig. 8. Restoration results of all methods on band 207 in Urban HSI data. (a) Original HSI. (b) SVD. (c) RegL1ALM. (d) CWM. (e) MoG-RPCA. (f) PMoEP. (g) LRMR. (h) LRTV. (i) TDL. (j) BM4D. (k) NMoG.

EO-1 Hyperion dataset.<sup>5</sup> The similar competing methods as the last section were compared in the real HSI experiments. Like the synthetic experiments, all involved parameters have been finely tuned or directly suggested by the original references to promise a possibly well performance of all competing methods. The component number of each band in the proposed method is set as 3 throughout all experiments. The gray value of HSIs were normalized into  $[0, 1]$ . The first Urban dataset is of the size  $307 \times 307 \times 210$ , and some bands are seriously polluted by atmosphere and water. We use all of data without removing any bands to more rationally verify the robustness of the proposed method in the presence of such heavy noises.

Figs. 7 and 8 present bands 103 and 207 of the Urban HSI obtained by all competing methods, respectively. From the figures, we can see that the original HSI bands are

contaminated by complex structural noise, like the stripe noise. It can be easily observed from Figs. 7(b)–(f) and 8(b)–(f) that the LRMF methods with different i.i.d. noise distribution assumptions cannot finely restore a clean HSI. This can be easily explained by the fact that such structural noises is obvious non-i.i.d. and the deviation between the real noise configuration and the encoded knowledge in the model then naturally degenerate the performance of the corresponding methods. Comparatively, albeit considering less prior knowledge on the to-be-recovered HSI, our method can still achieve a better recovery effect in visualization in its better restoration of texture and edge details and less preservation of structural noises due to its powerful noise modeling ability. This further substantiates the robustness of the proposed method in practical scenarios.

Then we give some quantitative comparison by showing the horizontal mean profiles of band 207 in Urban dataset

<sup>5</sup><http://datamirror.csdb.cn/admin/dataEO1Main.jsp>

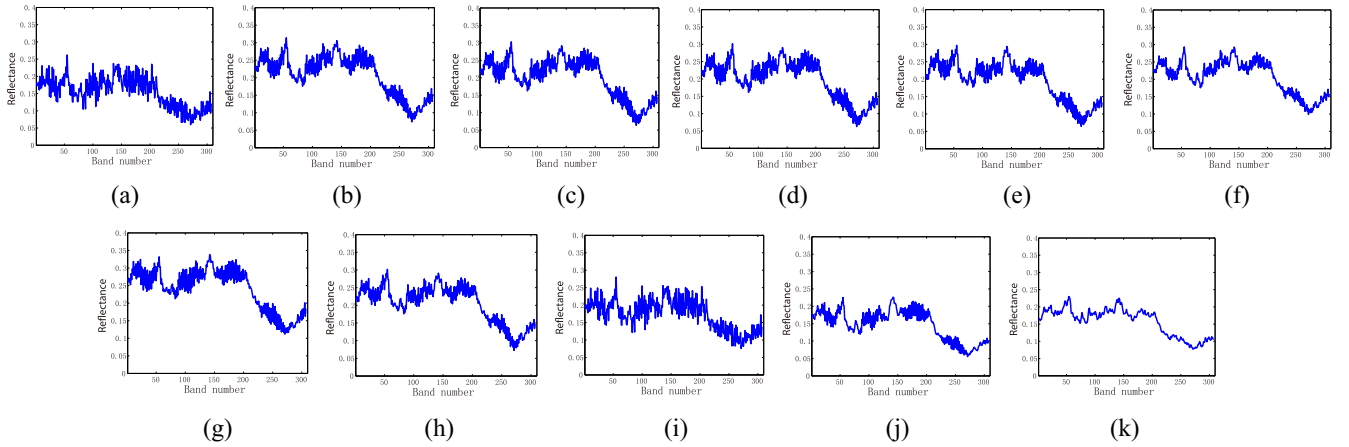


Fig. 9. Horizontal mean profiles of band 207 in Urban HSI data. (a) Original HSI. (b) SVD. (c) RegL1ALM. (d) CWM. (e) MoG-RPCA. (f) PMoEP. (g) LRMR. (h) LRTV. (i) TDL. (j) BM4D. (k) NMoG.

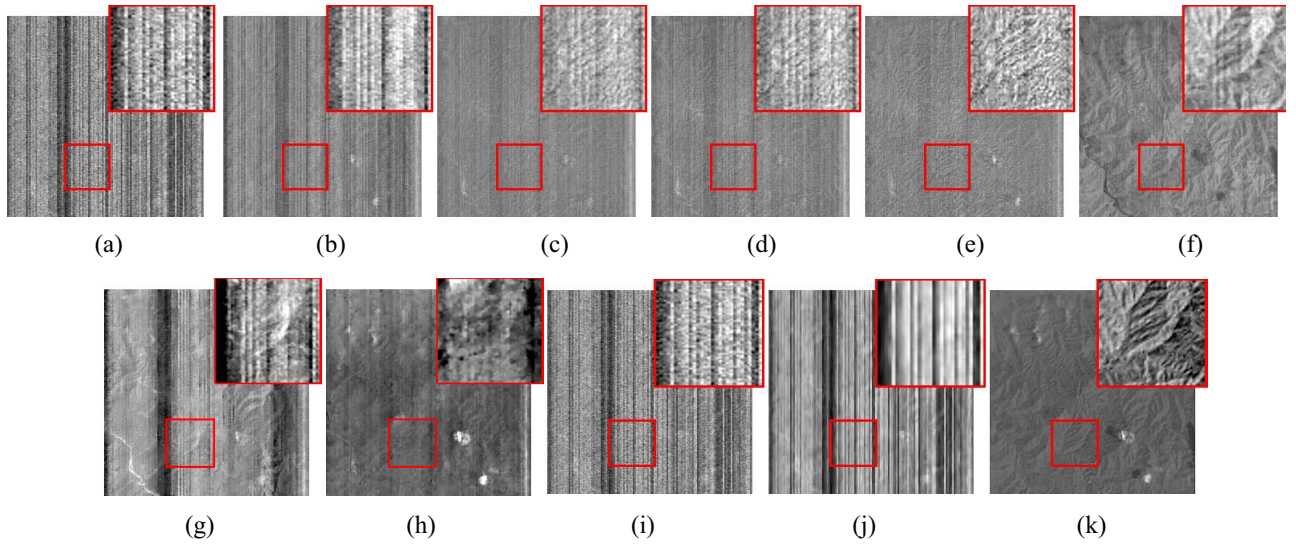


Fig. 10. Restoration results of all methods on band 100 in EO-1 Hyperion data. (a) Original HSI. (b) SVD. (c) RegL1ALM. (d) CWM. (e) MoG-RPCA. (f) PMoEP. (g) LRMR. (h) LRTV. (i) TDL. (j) BM4D. (k) NMoG.

before and after restoration in Fig. 9. The horizontal axis in the figure represents the row number, and the vertical axis represents the mean digital number value of each row. As shown in Fig. 9(a), due to the existence of mixed noise, there are rapid fluctuations in the curve. After the restoration processing, the fluctuations are more or less suppressed. It is easy to observe that the restoration curve of our NMoG-LRMF method provides evidently smoother curves.

The second Hyperion dataset is of the size  $3128 \times 256 \times 242$  and some bands are so seriously polluted that all signatures of pixel is equal 0. Thus we only use a subset of its 198 bands in this experiment after removing those zero signatures bands 1–7, 58–76, and 225–242, which are evidently outliers without any intrinsically useful information. We further spatially cropped a square area from the HSI with size  $256 \times 256 \times 198$  with relatively evident noises to specifically testify the denoising capability of a utilized method.

Figs. 10 and 11 show bands 100 and 197 of the restored HSI images by all competing methods. From the figures, it

can be easily observed that SVD and TDL methods have little effects in removing the noise; RegL1ALM and CWM methods can only partially eliminate the noise and MoG-RPCA, PMoEP, LRMR, and LRTV methods have a better performance than SVD, RegL1ALM, and CWM due to their better noise fitting ability brought by more appropriate models. They, however, still miss lots of details in their HSI recovery due to their implicit improper i.i.d. assumptions on noise distributions under HSI. All other competing methods also cannot achieve a satisfactory HSI denoising results even though they have considered more HSI prior knowledge in their models. Comparatively, the superiority of the proposed method can be easily observed in both detail preservation and noise removing.

Fig. 12 shows the vertical mean profiles of band 100 before and after restoration. As shown in Fig. 12(a), due to the existence of mixed noise especially the stripes disorderly located in the image, there are evident fluctuations over various places of the curve. After the restoration processing, all competing



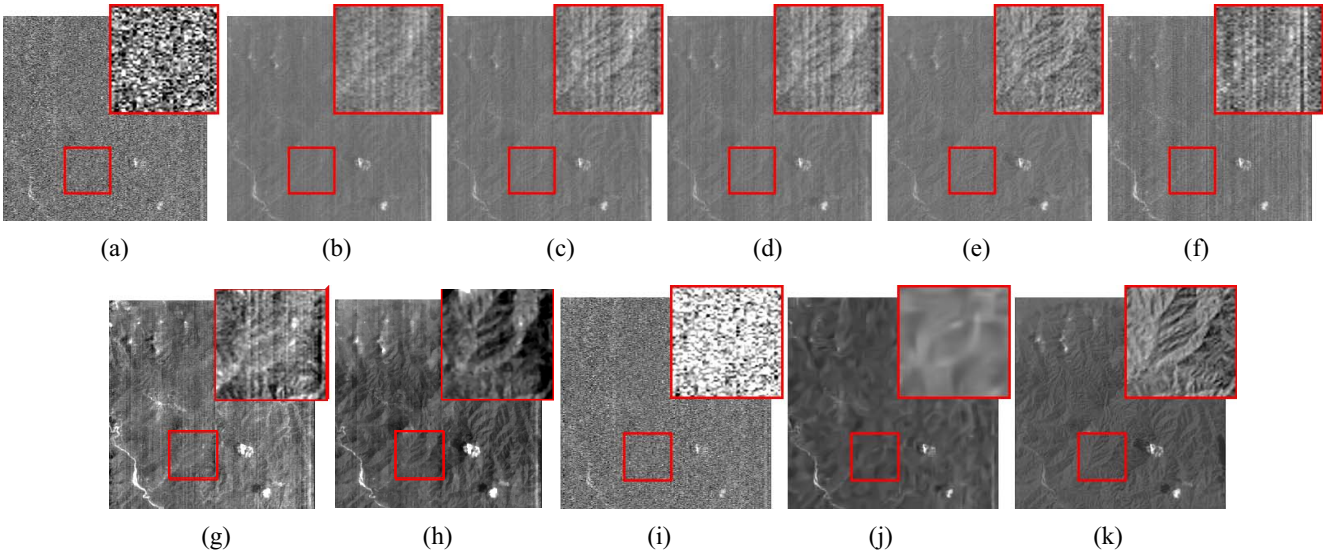


Fig. 11. Restoration results of all methods on band 197 in EO-1 Hyperion data. (a) Original HSI. (b) SVD. (c) RegL1ALM. (d) CWM. (e) MoG-RPCA. (f) PMoEP. (g) LRMF. (h) LRTV. (i) TDL. (j) BM4D. (k) NMoG.

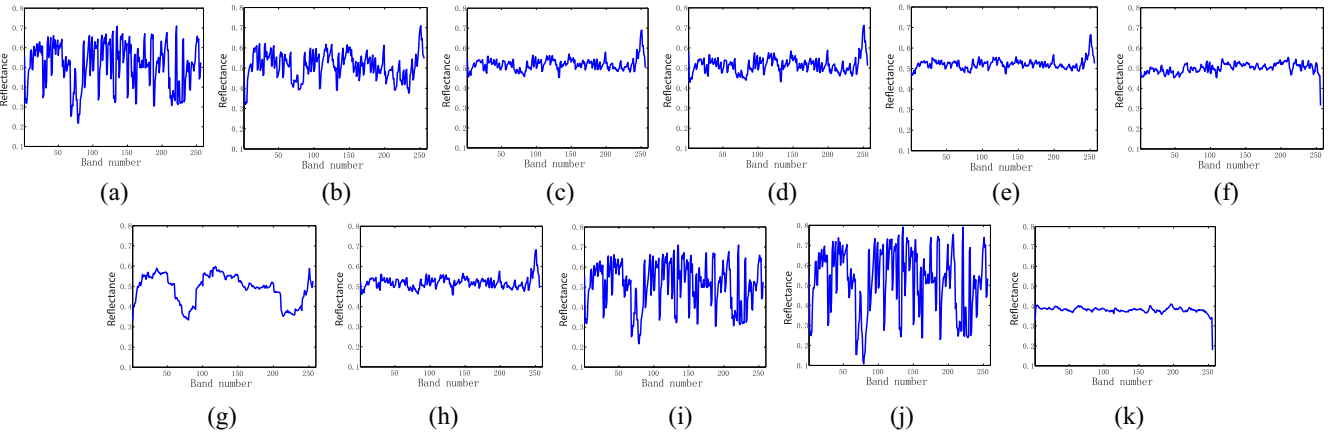


Fig. 12. Vertical mean profiles of band 100 in EO-1 Hyperion data. (a) Original HSI. (b) SVD. (c) RegL1ALM. (d) CWM. (e) MoG-RPCA. (f) PMoEP. (g) LRMF. (h) LRTV. (i) TDL. (j) BM4D. (k) NMoG.

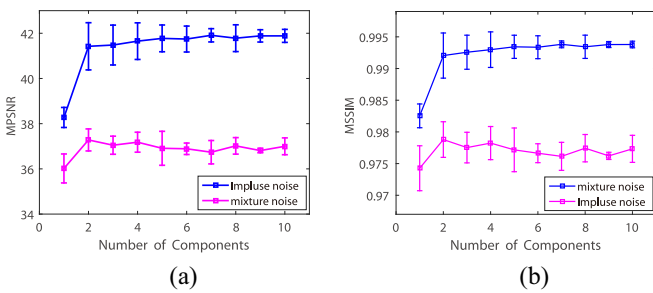


Fig. 13. Stability test of the proposed method to Gaussian component number  $K$  on DC Mall data in Gaussian + impulse noise and mixture noise cases. (a) MPSNR tendency curve with respect to  $K$ . (b) MSSIM tendency curve with respect to  $K$ .

methods show evident nonsmoothness over the corresponding curves, except that obtained from the proposed one. This implies that the HSI obtained by proposed NMoG-LRMF method can better preserve such prior knowledge possessed by real HSIs, as also substantiated by Fig. 9.

### C. Effect of Component Number on Denoising Performance

In this section, we examine the sensitivity of NMoG-LRMF method to the setting of the Gaussian component number  $K$ . We run NMoG-LRMF with 20 initializations on the DC Mall data in Gaussian + impulse noise and mixture noise cases, respectively, with  $K$  varying from 1 to 10. The results are shown in Fig. 13. It can be easily observed that after  $K$  is larger than 2, the denoising performance of the proposed method tends to be stable and not very sensitive to the choice of  $K$  value. Actually, in all our real experiments, we just simply set  $K$  as 3, and our method can consistently perform well throughout all our experiments.

## IV. CONCLUSION

In this paper, we initially propose a strategy to model the HSI noise using a non-i.i.d. noise assumption. Then we embed such noise modeling strategy into the LRMF model and propose a non-i.i.d LRMF model under the Bayesian framework. A VB algorithm is presented to infer the posterior of the



proposed model. Compared with the current state-of-the-arts techniques, the proposed method performs more robust due to its capability of adapting to various noise shapes encountered in applications, which is substantiated by our experiments implemented on synthetic and real noisy HSIs.

In our future work, we will try to extend the application of our method to video and face image data. Also, we will integrate more useful HSI prior terms into our model to further enhance its denoising capability. Besides, the proposed noise modeling strategy can be specifically redesigned under certain application context, like the wind speed prediction problem as indicated in [21] and [22].

## REFERENCES

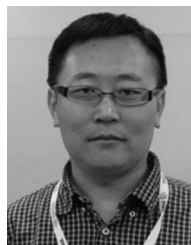
- [1] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
- [2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [4] X. Cao *et al.*, "Low-rank matrix factorization under general mixture noise distributions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, 2015, pp. 1493–1501.
- [5] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4677–4690, Oct. 2017.
- [6] G. Chen and S.-E. Qian, "Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis," *Can. J. Remote Sens.*, vol. 34, no. 5, pp. 447–454, 2008.
- [7] 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.
- [8] G. Chen and W.-P. Zhu, "Signal denoising using neighbouring dual-tree complex wavelet coefficients," *IET Signal Process.*, vol. 6, no. 2, pp. 143–147, Apr. 2012.
- [9] P. Chen, N. Wang, N. L. Zhang, and D.-Y. Yeung, "Bayesian adaptive matrix factorization with automatic model selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1284–1292.
- [10] S.-L. Chen, X.-Y. Hu, and S.-L. Peng, "Hyperspectral imagery denoising using a spatial-spectral domain mixing prior," *Comput. Sci. Technol.*, vol. 27, no. 4, pp. 851–861, 2012.
- [11] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. IEEE Conf. Signal Process.*, Poznań, Poland, 2007, pp. 145–149.
- [12] 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.
- [13] C. Ding, D. Zhou, X. He, and H. Zha, " $R_1$ -PCA: Rotational invariant  $L_1$ -norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, 2006, pp. 281–288.
- [14] 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.
- [15] A. Eriksson and A. Van Den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the  $L_1$  norm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 771–778.
- [16] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, Sep. 2009.
- [17] 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.
- [18] M. Haardt and P. Strobach, "Method for high-resolution spectral analysis in multi channel observations using a singular value decomposition (SVD) matrix technique," U.S. Patent 5 560 367, Oct. 1, 1996.
- [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] 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.
- [21] Q. Hu, Y. Wang, Z. Xie, P. Zhu, and D. Yu, "On estimating uncertainty of wind energy with mixture of distributions," *Energy*, vol. 112, pp. 935–962, Oct. 2016.
- [22] Q. Hu, S. Zhang, Z. Xie, J. Mi, and J. Wan, "Noise model based  $v$ -support vector regression with its application to short-term wind speed forecasting," *Neural Netw.*, vol. 57, pp. 1–11, Sep. 2014.
- [23] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
- [24] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 1791–1798.
- [25] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel non-negative Tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.
- [26] Q. Ke and T. Kanade, "Robust  $L_1$  norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, San Diego, CA, USA, 2005, pp. 739–746.
- [27] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.
- [28] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
- [29] L. Liu, L. Chen, C. L. P. Chen, Y. Y. Tang, and C. M. Pun, "Weighted joint sparse representation for removing mixed noise in image," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 600–611, Mar. 2017.
- [30] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.
- [31] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.
- [32] V. Maz'ya and G. Schmidt, "On approximate approximations using Gaussian kernels," *IMA J. Numer. Anal.*, vol. 16, no. 1, pp. 13–29, 1996.
- [33] D. Meng and F. De la Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, 2013, pp. 1337–1344.
- [34] D. Meng, Z. Xu, L. Zhang, and J. Zhao, "A cyclic weighted median method for  $L_1$  low-rank matrix factorization with missing entries," in *Proc. Assoc. Adv. Artif. Intell.*, vol. 4, Bellevue, WA, USA, 2013, p. 6.
- [35] K. Mitra, S. Sheorey, and R. Chellappa, "Large-scale matrix factorization with missing data under additional constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2010, pp. 1651–1659.
- [36] T. Okatani, T. Yoshida, and K. Deguchi, "Efficient algorithm for low-rank matrix factorization with missing components and performance comparison of latest algorithms," in *Proc. 28th Int. Conf. Mach. Learn.*, Barcelona, Spain, 2011, pp. 842–849.
- [37] 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.
- [38] Y. Peng *et al.*, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 2949–2956.
- [39] P. Sathya and K. Vani, "Hyperspectral image classification," *Adv. Nat. Appl. Sci.*, vol. 9, no. 6, pp. 338–344, 2015.
- [40] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, Jul. 2014.
- [41] D. W. J. Stein *et al.*, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

- [42] M. Wang, J. Yu, J.-H. Xue, and W. Sun, "Denoising of hyperspectral images using group low-rank representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4420–4427, Sep. 2016.
- [43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [44] X. Chen *et al.*, "Robust tensor factorization with unknown noise," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2016, pp. 5213–5221.
- [45] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2016, pp. 1692–1700.
- [46] R. Yan, L. Shao, and Y. Liu, "Nonlocal hierarchical dictionary learning using wavelets for image denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4689–4698, Dec. 2013.
- [47] 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.
- [48] A. C. Zelinski and V. K. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp.*, Denver, CO, USA, 2006, pp. 387–390.
- [49] 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.
- [50] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *Proc. 31st Int. Conf. Mach. Learn.*, Beijing, China, 2014, pp. 55–63.
- [51] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust L1-norm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, 2012, pp. 1410–1417.
- [52] P. Zhong and R. Wang, "Multiple-spectral-band CRFs for denoising junk bands of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2260–2275, Apr. 2013.



**Qian Zhao** received the B.Sc. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2009 and 2015, respectively.

He was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, USA, from 2014 to 2015. He is currently a Lecturer with the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include low-matrix/tensor analysis, Bayesian modeling, and self-paced learning.



**Deyu Meng** (M'13) received the B.Sc., M.Sc., and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively.

He is currently an Associate Professor with the Institute for Information and System Sciences, School of Mathematics and Statistics, Xi'an Jiaotong University. From 2012 to 2014, he took his two-year sabbatical leave with Carnegie Mellon University, Pittsburgh, PA, USA. His current research interests include self-paced learning, noise modeling, and tensor sparsity.



**Yang Chen** received the B.Sc. degree from Shaanxi Normal University, Xi'an, China, in 2013. She is currently pursuing the Ph.D. degree with Xi'an Jiaotong University, Xi'an.

Her current research interests include machine learning, noise modeling, and remote sensing image processing.



**Xiangyong Cao** received the B.Sc. degree from Xi'an Jiaotong University, Xi'an, China, in 2012, where he is currently pursuing the Ph.D. degree.

His current research interests include low-rank modeling, statistical modeling, and hyperspectral image analysis.



**Zongben Xu** received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987.

He serves as the Chief Scientist of the National Basic Research Program of China (973 Project), and the Director of the Institute for Information and System Sciences with Xi'an Jiaotong University. His current research interests include intelligent information processing and applied mathematics.

Dr. Xu was a recipient of the National Natural Science Award of China in 2007 and the CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a speech at the International Congress of Mathematicians 2010. He was elected as a member of the Chinese Academy of Science in 2011.