Adaptive Hyperspectral Mixed Noise Removal

Conference Paper · July 2018

DOI: 10.1109/IGARSS.2018.8519303

Tai-Xiang Jiang
Southwest University of Finance & Economics

Lina Zhuang
The University of Hong Kong

Ting-Zhu Huang
University of Electronic Science and Technology of China

ADAPTIVE HYPERSPECTRAL MIXED NOISE REMOVAL

Tai-Xiang Jiang[†,∗], *Lina Zhuang*[∗], *Ting-Zhu Huang*[†], *José M. Bioucas-Dias*[∗]

[†]School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, P. R. China

[∗]Instituto de Telecomunicaçõoes, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

ABSTRACT

This paper proposes a new denoising method for hyperspectral images (HSIs) corrupted by mixtures (in a statistical sense) of stripe noise, Gaussian noise, and impulsive noise. The proposed method has three distinctive features: **1**) it exploits the intrinsic characteristics of HSIs, namely, low-rank and selfsimilarity; **2**) the observation noise is assumed to be additive and modeled by a mixture of Gaussian (MoG) densities; **3**) the inference is performed with an expectation maximization (EM) algorithm, which, in addition to the clean HSI, also estimates the mixture parameters (posterior probability of each mode and variances). Comparisons of the proposed method with state-of-the-art algorithms provide experimental evidence of the effectiveness of the proposed denoising algorithm.

***Index Terms*—** Denoising, mixed noise, hyperspectral images, low-rank, selfsimilarity, mixture of Gaussians, expectation maximization.

1. INTRODUCTION

Hyperspectral remote sensing images are often unavoidably corrupted by several types of noises, including Gaussian noise, impulse noise, deadlines, and stripes [1]. Large research efforts have been devoted to HSI denoising. Among them, two critical points have been taken into consideration. One is that of preserving the structure of the clean HSIs while denoising. Another one is an appropriate modeling strategy for the noise.

As for HSI structure, HSIs are strongly correlated in the spectral-spatial domain, implying that they are low-rank, piecewise smooth, and selfsimilar. Low-rank has been exploited by representing the spectral vectors in low-dimensional subspaces [1–4], low-rank matrix recovery (LRMR) [5], noise adjusted iterative low-rank matrix approximation (NAILRMA) [6], and nonconvex low-rank matrix approximation (NonLRMA) [7]. Piecewise smoothness in the spatial domain has been exploited, for example, via total-variation regurarization (see, e.g., [8], [4]).

Meanwhile, image selfsimilarity underlies the state of the art in gray-level image denoising. This form of prior, or regularizer, has been fully exploited in nonlocal means [9], BM3D [10] and LRCF [11]. Zhuang et al. proposed a series of methods (FastHyDe [2], RHyDe [3] and GLF [12]), which tactfully exploit HSI low-rank and selfsimilarity. In short, these methods start by identifying the subspace where the spectral vectors live, and then formulate the denoising problem with respect to the representation coefficients in the subspace, which are also selfsimilar.

Although FastHyDe [2] and GLF [12] achieved unexceptionable results for the task of Gaussian noise and Poisson noise removal, they are not robust to the mixed noise. To cope with this issue, RHyDe [3] took account of the dead pixels by decomposing the noise into a Gaussian term plus a sparse term and imposing mixed $\ell_{2,1}$ regularization on the latter term. LRMR [5], NAILRMA [6], NonLRMA [7], and LRTV [8] also adopt sparsity inducing regularization term, but contrarily to RHyDe, formulate the inference in the original HSI domain. As mentioned before, in many real applications, the noise often exhibits very complex statistical distributions. This motivates us to consider a more flexible modeling strategy to tackle such complex noise cases.

Contribution We model the noise as an additive term with a MoG density, which is a universal approximation to any continuous distribution and hence capable of modeling a wider range of noise distributions. To automatically estimate the parameters involved in different noise distributions, we design an expectation maximization (EM) algorithm. Similar to [2, 3, 12], our method takes full advantage of the spectral low dimension and spatial self-similarity of the HSIs.

The outline of this paper is given as follows. Section 2.1 formulates the problem of HSI denoising. Section 2.2 gives the problem formulation and the proposed EM algorithm. Experimental results including comparisons are reported in Section 3. Finally, we draw some conclusions in Section 4.

2. MAIN RESULTS

Let $\mathbf{Y} \in \mathbb{R}^{b\times n}$ denote a HSI with spatial resolution $n = w \times h$ (spatial width $\times$ spatial height) and b spectral bands. The noise is assumed to be additive. Therefore, we may write

This research was supported by the National Natural Science Foundation of China (61772003). The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7-PEOPLE-2013-ITN) under grant agreement n607290 SpaRTaN. This work was partially supported by the Fundaçáo para a Ciência e Tecnologia, Portuguese Ministry of Science and Higher Education, projects UID/EEA/50008/2013 and ERANETMED/0001/2014.

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

where, $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{N} \in \mathbb{R}^{b\times n}$ are, respectively, the observed HSI, the clean HSI, and the noise.

Since, with good approximation, we are assuming that the spectral vectors (columns of $\mathbf{X}$) live in a subspace, we may write [1, 2]

$$\mathbf{Y} = \mathbf{E}\mathbf{Z} + \mathbf{N}, \tag{2}$$

where $\mathbf{E} \in \mathbb{R}^{b\times s}$, with $b \ll s$, and $\mathbf{Z} \in \mathbb{R}^{s\times n}$. Matrix $\mathbf{E}$, assumed to be semi-unitary (i.e., $\mathbf{E}^T\mathbf{E} = \mathbf{I}$), spans the subspace. The subspace is estimated with, for example, the HySime algorithm [13]. Matrix $\mathbf{Z}$ contains the representation coefficients of $\mathbf{X}$ with respect to $\mathbf{E}$.

2.1. Problem Formulation

Using Bayes rule, the posteriori probability distribution of $\mathbf{Z}$ conditioned to $\mathbf{Y}$ is given by

$$p(\mathbf{Z}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{Z})p(\mathbf{Z})}{p(\mathbf{Y})}, \tag{3}$$

where $p(\mathbf{Y}|\mathbf{Z})$ is the probability of $\mathbf{Y}$ given $\mathbf{Z}$ (the *likelihood function*) and $p(\mathbf{Z})$ is *a priori* probability density function of $\mathbf{Z}$. The *maximum a posteriori* (MAP) estimate of $\mathbf{Z}$ is

$$\widehat{\mathbf{Z}} \in \arg\max_{\mathbf{Z}} \ln p(\mathbf{Y}|\mathbf{Z}) + \ln p(\mathbf{Z}). \tag{4}$$

In this paper, we consider that the noise is a MoG mixture with only two modes: the first mode models i.i.d. zero-mean white Gaussian noise with variance $\sigma_{i,1}^2$ for the i-th band. The second term models stripe and impulsive noise in the i-th band and it is assumed to follow a Gaussian distribution with zero-mean and a very large variance $\sigma_{i,2}^2$ ($\sigma_{i,2}^2 \gg \sigma_{i,k}$). Hence,

$$p(y_{ij}|x_{ij}) = \sum_{k=1}^{2} \alpha_k \mathcal{N}(y_{ij} - x_{ij}, \sigma_{i,k}^2), \tag{5}$$

where $y_{ij} := [\mathbf{Y}]_{ij}$, $x_{ij} := [\mathbf{E}\mathbf{Z}]_{ij}$, $\alpha_k \geq 0$ is the probability of mode $k \in \{1, 2\}$, and $\mathcal{N}(y - \mu, \sigma^2)$ denotes a Gaussian density with mean μ and variance σ^2 computed at y.

Assuming that $\mathbf{Z}$ and $\mathbf{N}$ are independent, it follows $p(\mathbf{Y}|\mathbf{Z}) = \prod_i \prod_j p(y_{ij}|x_{ij})$. Then, the MAP problem (4) turns out to be

$$\widehat{\mathbf{\Theta}} \in \arg\max_{\mathbf{\Theta}} \left\{ \prod_{i=1}^{b} \prod_{j=1}^{n} \left(\sum_{k=1}^{2} \alpha_k \mathcal{N}(y_{ij} - x_{ij}, \sigma_k^2) \right) p(\mathbf{Z}) \right\}, \tag{6}$$

where $\mathbf{\Theta} := \{\mathbf{Z}, \alpha_k, \sigma_{i,k}^2\}$ ($i = 1, ..., b$ and $k = 1, 2$).

2.2. Proposed EM algorithm

Problem (6) is nonconvex and we use the EM algorithm [14] to compute a local optima. To apply the EM algorithm, as usual in mixtures, we introduce the latent variables u_{ijk} (playing the role of missing data), for $ijk \in \{1, \ldots, b\} \times \{1, \ldots, n\} \times \{1, 2\}$, which selects the active mode at band i and pixel j.

Let $\mathbf{\Theta}^{(t)} := \{\mathbf{Z}^{(t)}, \alpha_k^{(t)}, (\sigma_{i,k}^2)^{(t)}\}$ ($i = 1, ..., b$ and $k = 1, 2$) denote the set of parameters at the t-th iteration of the EM algorithm. Then, the E-step and the M-step amounts to compute (see [14] for details)

E-Step

$$\omega_{ij,k}^t = \mathbb{E}[u_{ij,k}|\mathbf{Y}, \mathbf{\Theta}^{(t)}] = \frac{\alpha_1^t \mathcal{N}\left(y_{ij} - x_{ij}^{(t)}, (\sigma_{i,k}^2)^{(t)}\right)}{\sum_{k=1}^{2} \alpha_k^t \mathcal{N}\left(y_{ij} - x_{ij}^{(t)}, (\sigma_{i,k}^2)^{(t)}\right)} \tag{7}$$

M-step Construct the so-called Q function:

$$\begin{aligned} Q\left(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}\right) = \sum_{i=1}^{b} \sum_{j=1}^{n} \sum_{k=1}^{2} \omega_{ij,k}^t \Bigg(-\frac{1}{2}\log(2\pi\sigma_{i,k}^2) \\ -\frac{(y_{ij} - x_{ij})^2}{2\sigma_{i,k}^2} + \log \alpha_{i,k} \Bigg) + \log p(\mathbf{Z}). \end{aligned} \tag{8}$$

Then, optimize $Q\left(\mathbf{\Theta}, \mathbf{\Theta}^{(t)}\right)$ with respect to $\mathbf{\Theta}$. The optimization w.r.t. $\alpha_{i,k}, \sigma_{i,k}$, for $k = 1, 2$ and $i = 1, ..., b$, yields

$$\alpha_{i,k} = \frac{\sum_j \omega_{ij,k}^t}{\sum_{j,p} \omega_{ij,p}^t}, \quad \sigma_{i,k}^2 = \frac{\sum_j \omega_{ij,k}^t (y_{ij} - x_{ij})^2}{\sum_{ij,p} \omega_{ij,k}^t}. \tag{9}$$

We remark that $\alpha_2 = 1 - \alpha_1$. The optimization w.r.t. $\mathbf{Z}$ is (recall that $x_{ij} = [\mathbf{E}\mathbf{Z}]_{ij}$)

$$\min_{\mathbf{Z}} \sum_{i=1}^{b} \sum_{j=1}^{n} \sum_{k=1}^{2} \omega_{ij,k}^t \left(\frac{(y_{ij} - x_{ij})^2}{2\sigma_{i,k}^2} \right) - \log p(\mathbf{Z}). \tag{10}$$

Optimization (10) may be compactly written as

$$\min_{\mathbf{Z}} \sum_{k=1}^{2} \frac{1}{2} \|\mathbf{M}_k \odot (\mathbf{Y} - \mathbf{E}\mathbf{Z})\|_F^2 + \log p(\mathbf{Z}), \tag{11}$$

where $[\mathbf{M}_k]_{i,j} := \sqrt{\omega_{ij,k}^t / \sigma_{i,k}^2}$, $\|\mathbf{X}\|_F^2 := \sqrt{\text{trace}(\mathbf{X}\mathbf{X}^\top)}$, and $\odot$ stands for elementwise multiplication. Considering that $\sigma_{i,2} \gg \sigma_{i,1}$, optimization (11) is well approximated by

$$\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{M}_1 \odot (\mathbf{Y} - \mathbf{E}\mathbf{Z})\|_F^2 + \lambda\phi(\mathbf{Z}), \tag{12}$$

where $\lambda\phi(\mathbf{Z}) := -\log p(\mathbf{Z})$ and $\lambda > 0$ acts as a regularization parameter. We note that (12) is a convex problem, provided that ϕ is convex.

We use SALSA [15] to solve (12). To set the stage for SALSA, we start by reformulating (12) as the equivalent constrained optimization

$$\begin{aligned} &\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{M}_k \odot (\mathbf{Y} - \mathbf{V}_1)\|_F^2 + \lambda\phi(\mathbf{V}_2) \\ &\text{s.t.} \mathbf{V}_1 = \mathbf{E}\mathbf{Z}, \ \mathbf{V}_2 = \mathbf{Z}. \end{aligned} \tag{13}$$

The augmented Lagrangian function for (13) is

$$\begin{aligned} L(\mathbf{Z}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{D}_1, \mathbf{D}_2) = \frac{1}{2} \|\mathbf{M}_1 \odot (\mathbf{Y} - \mathbf{V}_1)\|_F^2 + \lambda\phi(\mathbf{V}_2) \\ + \frac{\mu}{2} \|\mathbf{E}\mathbf{Z} - \mathbf{V}_1 - \mathbf{D}_1\|_F^2 + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{V}_2 - \mathbf{D}_2\|_F^2, \end{aligned} \tag{14}$$

where $\mathbf{D}_1, \mathbf{D}_2$ are scaled Lagrangian multipliers and $\mu > 0$. Then SALSA iteratively optimize (14) w.r.t $\mathbf{Z}, \mathbf{V}_1, \mathbf{V}_2$ and update $\mathbf{D}_1, \mathbf{D}_2$, leading to the following updates:

$$\begin{cases} \mathbf{Z}^{(p+1)} = \frac{1}{2}\left(\mathbf{E}^{\top}(\mathbf{V}_1^{(p)} + \mathbf{D}_1^{(p)}) + \mathbf{D}_2^{(p)} + \mathbf{V}2^{(p)}\right), \\ \mathbf{V}_1{}^{(p+1)} = (\mathbf{E}\mathbf{Z}^{(p+1)} - \mathbf{D}_1^{(p)} + \mathbf{M}_1 \odot \mathbf{M}_1 \odot \mathbf{Y}) \\ \qquad \oslash (\mathbf{M}_1 \odot \mathbf{M}_1 \oplus \mu) \\ \mathbf{V}_2{}^{p+1} = \arg\min_{\mathbf{V}_2} \frac{\mu}{2}\|\mathbf{Z}^{(p+1)} - \mathbf{V}_2 - \mathbf{D}_2^{(p)}\|_F^2 + \lambda\phi(\mathbf{V}_2) \\ \mathbf{D}_1{}^{(p+1)} = \mathbf{D}_1^{(p)} - (\mathbf{E}\mathbf{Z}^{(p+1)} - \mathbf{V}_1^{(p+1)}) \\ \mathbf{D}_2{}^{(p+1)} = \mathbf{D}_2^{(p)} - (\mathbf{Z}^{(p+1)} - \mathbf{V}_2^{(p+1)}). \end{cases} \tag{15}$$

To solve the $\mathbf{V}_2$ subproblem, we resort to the plug-and-play (PnP) prior framework [16]. As in [2] we use BM3D, which is a very fast state-of-the-art denoiser conceived to enforce self-similarity.

Pre-processing The EM algorithm is initialized with $\mathbf{\Theta}^{(0)} = \{\mathbf{Z}^{(0)}, \alpha_k^{(0)}, (\sigma_{i,k}^2)^{(0)}\}$, where $\mathbf{Z}^{(0)} = \mathbf{E}^T\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}$ is obtained by bandwise pre-filtering the noisy HSI with a 3×3 median filter, $\alpha_k^{(0)}$ and $(\sigma_k^2)^{(0)}$ given by (9),

$$\omega_{ij,1}^{(0)} = \begin{cases} 1, & \text{if } |\widetilde{x}_{ij} - y_{ij}| < 3\widehat{\sigma}_{i,1}, \\ 0, & \text{otherwise}, \end{cases} \quad \omega_{ij,2}^{(0)} = 1 - \omega_{ij,1}^{(0)}, \tag{16}$$

with $\widetilde{\mathbf{X}} = \mathbf{E}\mathbf{Z}^{(0)}$, $\widetilde{x}_{ij} := [\widetilde{\mathbf{X}}]_{ij}$, and $\widehat{\sigma}_{i,1}$ given by the sample variance of the vector $\widetilde{\mathbf{Y}}(i,:) - \widetilde{\mathbf{X}}(i,:)$.

Algorithm 1 EM algorithm for HSIs denoising

Input: $\mathbf{Y} \in \mathbb{R}^{b \times n}$

Pre-processing: $\widetilde{\mathbf{Y}} = \text{med}(\mathbf{Y})$; $\mathbf{E} = \text{HySime}(\widetilde{\mathbf{Y}})$; set $\omega_{ij,k}^{(0)}$, $\alpha_k^{(0)}$, $(\sigma_{i,k}^2)^{(0)}$ using (16) and (9).

repeat
    (E-step): Update $\omega_{ij,k}^{(t)}$ via (7)
    (M-step):
        Update $\alpha_{i,k}^{(t)}$ and $(\sigma_{i,k}^2)^{(t)}$ via (9)
        Update $\mathbf{Z}^{(t)}$ by running a number of SALSA iterations (15)
        $\mathbf{X}^{(t)} = \mathbf{E}\mathbf{Z}^{(t)}$
until converge;

Output: The denoised HSI $\mathbf{X}$

Algorithm 1 shows the pseudocode for the proposed HSI denoising method.

3. EVALUATION WITH SIMULATED DATA

To compare the proposed method with the state-of-the-art denoising algorithms, we conduct experiments using a subimage of Washington DC Mall dataset[1] (of size $n = 256 \times 256$, $b = 191$) and Pavia city center dataset[2] (of size $n = 610 \times 339$, $b = 87$). The bands of the clean images were normalized to $[0, 1]$. These subimages of high quality are considered as the clean HSIs. Two noisy datasets were generated as follows.

Case 1 Synthetic data with Gaussian noise, impulsive noise, and stripe noise. Gaussian noise is i.i.d. and has zero-mean and $\sigma^2 = 0.1^2$. The impulsive noise (salt and pepper) is added to all bands of the HSI. The impulsive noise affects 10% of the pixels. The stripe noise is made of vertical tripes affecting 10% bands and, for each band, 20% of the pixels. The width of the stripes varies from one line to three lines.

Case 2 Synthetic data with Gaussian noise and stripe noise. The Gaussian noise is bandwise i.i.d. with zero-mean and variance σ_i^2, for $i = 1, \ldots, b$, sampled from a Uniform distribution $U(0, 0.1)$. The stripe noise with different shapes (vertical lines, oblique lines, curves) affect 30% bands of the bands and, for each band, about 10% of the pixels.

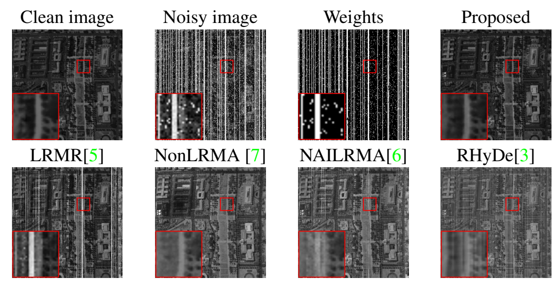

Fig. 1. Denoising results produced by different algorithms for band 183 of Washington DC Mall. 1st row from left to right: clean image, noisy image, weights corresponding to the sparse noise including stripes, and denoising result by the proposed method. 2nd row: denoising results for different algorithms.

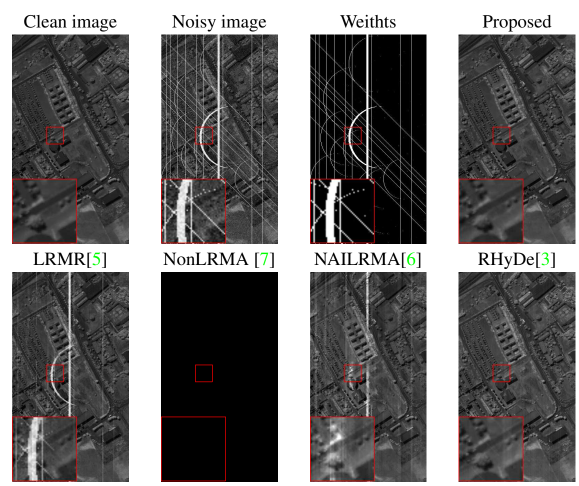

Fig. 2. Denoising results for band 81 of Pavia city center. 1st row from left to right: clean image, noisy image, weights corresponding to the sparse noise including stripes, and the denoising result by the proposed method. 2nd row: denoising results from different algorithms.

The state-of-the-art hyperspectral denoising methods LRMR[5], NonLRMA [7], NAILRMA [6], and RHyDe [3] are uses for comparison. The mean values of peak signal-to-noise (PSNR) index and the structural similarity (SSIM) index of each band are calculated for quantitative assessment and reported in Table 1. It can be seen from Table 1 that the proposed method outperforms the state-of-the-art methods, dealing with different types of mixed noise.

Figure 1-2 illustrate the denoising results and the obtained weights corresponding to the sparse noise including stripes,

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

Table 1. Quantitative comparisons

	Data	Index	Noisy image	LRMR[5]	NonLRMA [7]	NAILRMA[6]	RHyDe[3]	Proposed
Case 1	Washington DC Mall	MPSNR	13.49	31.74	22.89	24.97	26.21	**33.24**
		MSSIM	0.2446	0.9089	0.7359	0.8157	0.8503	**0.9566**
		running time (s)	–	105.9	1044.7	142.2	**51.6**	351.4
Case 2	Pavia city center	MPSNR	23.08	30.12	16.35	30.53	27.61	**33.63**
		MSSIM	0.4686	0.8564	0.0090	0.8774	0.7184	**0.9372**
		running time (s)	–	180.3	205.7	390.2	**96.64**	1303.8

for the DC Mall data with noise case 1 and the Pavia city center data with noise case 2, respectively. In either situation, the noisy image is seriously degraded for the mixed noise. The results recovered by LRMR and NAILRMA achieve comparatively high MPSNR but are visually undesirable. NonLRMA yields a distortion of pixel intensities for case 1 and fails for case 2. Meanwhile, the stripes affect the estimation of subspace by RhyDe, which uses the singular value decomposition. The proposed method removes almost all the noise and preserves geometric features and details. The estimated weights for the sparse noise including stripes accurately accord with the noise's location. It is notable in Figure 3 that for the noise case 2 the proposed method well estimated the bandwisely different variances of the Gaussian noise.

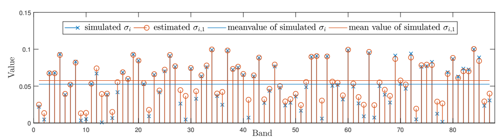

Fig. 3. The values of simulated σ_i and estimated $\sigma_{i,1}$ with respect to band numbers, for noise case 1.

4. CONCLUSION

This paper introduces a new HSI denoising tailored to mixtures of Gaussian noise, impulsive noise, and stipe noise. On one hand, the proposed method simultaneously exploits two intrinsic characteristics of HSIs, i.e, the high correlation along the spectral mode and the nonlocal similarity along the spatial modes. On the other hand, a MoG is used to model the mixed noise and the distribution of different types of noise is estimated including their locations. In this sense, the proposed denoiser automatically adapts to the noise statistics. A limited comparison of the proposed method with the state-of-the-art algorithms is conducted. The results on the simulated data show the superiority of the proposed method for complex noise.

References

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 5, no. 2, pp. 354–379, 2012.

[2] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 730–742, 2018.

[3] L. Zhuang, L. Gao, B. Zhang, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," in *SPIE - Conf. on Image and Signal Processing for Remote Sensing*, vol. 10427, September 2017, pp. 10 427–10 427–8.

[4] M. Simões, J. Bioucas-Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[5] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.

[6] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3050–3061, 2015.

[7] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5366–5380, 2017.

[8] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 178–188, 2016.

[9] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 60–65.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[11] M. Nejati, S. Samavi, S. R. Soroushmehr, and K. Najarian, "Low-rank regularized collaborative filtering for image denoising," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 730–734.

[12] L. Zhuang and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and non-local low-rank factorizations," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 1900–1904.

[13] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[15] M. V. Afonso, J. M. Bioucas-Dias, and M. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.

[16] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2013, pp. 945–948.