Self-Paced Nonnegative Matrix Factorization for Hyperspectral Unmixing

Jiangtao Peng, Yicong Zhou, *Senior Member, IEEE*, Weiwei Sun, *Member, IEEE*, Qian Du, *Fellow, IEEE*, and Lekang Xia

***Abstract*— The presence of mixed pixels in the hyperspectral data makes unmixing to be a key step for many applications. Unsupervised unmixing needs to estimate the number of endmembers, their spectral signatures, and their abundances at each pixel. Since both endmember and abundance matrices are unknown, unsupervised unmixing can be considered as a blind source separation problem and can be solved by nonnegative matrix factorization (NMF). However, most of the existing NMF unmixing methods use a least-squares objective function that is sensitive to the noise and outliers. To deal with different types of noises in hyperspectral data, such as the noise in different bands (band noise), the noise in different pixels (pixel noise), and the noise in different elements of hyperspectral data matrix (element noise), we propose three self-paced learning based NMF (SpNMF) unmixing models in this article. The SpNMF models replace the least-squares loss in the standard NMF model with weighted least-squares losses and adopt a self-paced learning (SPL) strategy to learn the weights adaptively. In each iteration of SPL, atoms (bands or pixels or elements) with weight zero are considered as complex atoms and are excluded, while atoms with nonzero weights are considered as easy atoms and are included in the current unmixing model. By gradually enlarging the size of the current model set, SpNMF can select atoms from easy to complex. Usually, noisy or outlying atoms are complex atoms that are excluded from the unmixing model. Thus, SpNMF models are robust to noise and outliers. Experimental results on the simulated and two real hyperspectral data sets demonstrate that our proposed SpNMF methods are more accurate and robust than the existing NMF methods, especially in the case of heavy noise.**

Manuscript received April 7, 2020; accepted May 18, 2020. Date of publication June 10, 2020; date of current version January 21, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61871177, Grant 41971296, Grant 11771130, and Grant 41671342, in part by the Science and Technology Development Fund, Macau SAR under Grant 189/2017/A3, in part by the University of Macau under Grant MYRG2018-00136-FST, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR19D010001, and in part by the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 18R05. *(Corresponding author: Weiwei Sun.)*

Jiangtao Peng and Lekang Xia are with the Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China (e-mail: pengjt1982@hubu.edu.cn).

Yicong Zhou is with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yicongzhou@um.edu.mo).

Weiwei Sun is with the Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China (e-mail: sunweiwei@nbu.edu.cn).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: du@ece.msstate.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2996688

***Index Terms*— Hyperspectral unmixing, nonnegative matrix factorization (NMF), self-paced learning (SPL).**

I. Introduction

HYPERSPECTRAL sensors can capture image data across hundreds of narrow and contiguous spectral bands. Due to the low spatial resolution of hyperspectral sensors, microscopic material mixing, and multiple scattering [1], each image pixel may cover several different materials, and its spectrum is a mixture of several pure materials' spectra. The existence of mixed spectra dramatically affects the application of hyperspectral data. A technique to deal with the mixed spectra problem is hyperspectral unmixing [1], which aims at decomposing a mixed spectrum into pure materials' spectra (endmembers) and their corresponding fractions (abundances). It is clear that two main tasks of hyperspectral unmixing are endmember extraction and abundance estimation. These two steps can be performed sequentially, i.e., first extracting endmembers by n-finder algorithm (N-FINDR) [2] or vertex component analysis (VCA) [3] and then estimating abundances by least-squares methods. However, endmember extraction methods in the first step often assume the presence of pure pixels in the data [2]–[4]. When pure pixels are absent, traditional endmember extraction methods will fail. Fortunately, blind source separation (BSS) algorithms, such as nonnegative matrix factorization (NMF) methods, can be used to unmix hyperspectral data without the assumption of pure pixels.

In the NMF, a matrix is factorized into two low-rank matrices with the property that all three matrices have no negative elements. The nonnegativity and part-based representation makes the resulting matrices easier to be interpreted and suitable for many real applications [5], such as clustering, feature extraction, and hyperspectral unmixing. For hyperspectral unmixing problem, NMF can decompose a nonnegative hyperspectral data matrix into a nonnegative endmember matrix and a nonnegative abundance matrix simultaneously. During past decades, many NMF-based unmixing methods have been proposed. Miao and Qi [6] proposed a minimum volume-constrained NMF (MVCNMF) method for unsupervised endmember extraction from highly mixed image data. Qian *et al.* [5] proposed an $\ell_{1/2}$-NMF method that imposed an $\ell_{1/2}$ sparsity constraint on the abundance matrix. Wang *et al.* [7] proposed an endmember dissimilarity-constrained NMF method. Feng *et al.* [8] developed a

sparsity-constrained deep NMF with total variation technique for hyperspectral unmixing. Du *et al.* [9] made breakthrough results that, for the first time, considered spectral unmixing as a multitask learning-based blind source separation problem and provided a promising solution to the automatic hyperspectral image processing problem.

The aforementioned NMF methods improved the original NMF unmixing method by imposing additional constraints either on the endmember matrix or on the abundance matrix. When there exists noise in the hyperspectral data, the performance of these models will dramatically degrade. Due to the imaging mechanism, the hyperspectral data inevitably include different types of noises [10], [11], i.e., Gaussian noise, impulse noise, stripes, and deadlines. To cope with noisy bands, Wang *et al.* [12] proposed a correntropy-based robust NMF (CENMF) method that employed a correntropy-based metric to replace the least-squares metric and used a half-quadratic technique to solve the model. Rather than using the half-quadratic technique, Zhu *et al.* [13] employed alternating direction method of multipliers (ADMM) technique to solve the correntropy-based NMF model. The noise sensitive least-squares metric in the NMF can also be replaced by other robust metrics, such as $\ell_{2,1}$-norm- and $\ell_{1,2}$-norm-based metrics [14], [15]. Huang *et al.* [4] combined the $\ell_{2,1}$-norm and $\ell_{1,2}$-norm and proposed a spectral–spatial robust NMF (SSRNMF) model to deal with the rowwise band noise and columnwise pixel noise.

To deal with the rowwise band noise, columnwise pixel noise, and elementwise noise, in this article, we employ self-paced learning (SPL) strategy to select some "good" atoms (i.e., bands or pixels or elements) for NMF unmixing and propose three self-paced NMF (SpNMF) models to handle the band noise, pixel noise, and element noise, respectively. Our SpNMF model can be considered as a regularized weighted NMF model that uses a self-paced regularizer to adaptively learn weight for each atom based on the approximation error and then uses the learned weights to eliminate the effect of noisy atoms. According to the nature of SPL, the weight vector can be used to indicate the importance of atoms, and only the important atoms with nonzero weights (i.e., "easy" or "good" atoms) are included in the current unmixing model. Thus, SpNMF methods always select some "good" atoms into the unmixing model and, thus, can automatically exclude "complex" or "bad" atoms. Due to the atom-selection ability of SPL, SpNMF models can provide robust performance in outlier or heavy noise cases [16], [17].

The contributions of this article are threefold.

1) We propose three robust SPL-based hyperspectral unmixing methods to handle band, pixel, and element noise called self-paced NMF for band weighting (SpNMFB), self-paced NMF for pixel weighting (SpNMFP), and self-paced NMF for element weighting (SpNMFE) and develop an alternative optimization strategy (AOS) to find corresponding solutions.
2) We provide a comparison of different NMF-based unmixing methods from the viewpoints of the loss function, regularizer, and weight and also explain that SpNMF methods are robust to noise or outliers.
3) We evaluate different NMF-based unmixing methods on both simulated and real data sets. Extensive experimental results validate the effectiveness and robustness of SpNMF methods for hyperspectral unmixing.

The rest of this article is organized as follows. In Section II, NMF unmixing models are reviewed. Section III describes the proposed self-paced NMF models. The experimental results and analysis are provided in Section IV. Finally, Section V draws the conclusion.

II. NMF Unmixing Model

Due to the limited spatial resolution, an HSI pixel usually covers several different materials, and its spectral response is a mixture of these materials [1], [5]. Under the linear mixing mechanism, each observed pixel $\mathbf{y} \in \mathcal{R}^{B\times 1}$ can be represented as a linear combination of several spectral signatures called endmembers, i.e., $\mathbf{x}_1, \ldots, \mathbf{x}_P$

$$\mathbf{y} = \mathbf{x}_1 w_1 + \cdots + \mathbf{x}_P w_P + \mathbf{e} = \mathbf{X}\mathbf{w} + \mathbf{e} \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P] \in \mathcal{R}^{B\times P}$ is a nonnegative spectral signature matrix, $\mathbf{w} = [w_1; \cdots; w_P] \in \mathcal{R}^{P\times 1}$ is the abundance fraction for each endmember, and $\mathbf{e}$ is the additive noise vector.

Assume that there are N pixels in the HSI. The linear mixing model (LMM) can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{W} + \mathbf{E} \tag{2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathcal{R}^{B\times N}$ is the hyperspectral data, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N] \in \mathcal{R}^{P\times N}$ is the endmember abundance matrix, and $\mathbf{E}$ is the noise matrix. In the LMM framework, the endmember and abundance matrices should be nonnegative from the viewpoint of physical meaningfulness. Meanwhile, the abundance vector for each pixel should satisfy the sum-to-one constraint [1].

For the unmixing problem, we only have the hyperspectral data $\mathbf{Y}$ and need to solve both the endmember and abundance matrices. Due to the nonnegative nature, the NMF model can be used for hyperspectral unmixing [5], [6], [18]. The NMF unmixing model can be written as

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{W}} \left\|\mathbf{Y} - \mathbf{X}\mathbf{W}\right\|_F^2, \\ &\text{s.t. } \mathbf{X}_{bp} \geq 0, \mathbf{W}_{pn} \geq 0 \quad \forall b, p, n \\ &\qquad \sum_{p=1}^{P} \mathbf{W}_{pn} = 1, \quad n = 1, \ldots, N \end{aligned} \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Due to nonnegative constraints on $\mathbf{X}$ and $\mathbf{W}$, NMF can learn a part-based representation of data, which has been proven to be suitable for hyperspectral unmixing [5], [18]. The NMF unmixing model (3) can be solved by the multiplicative update algorithm [19], [20]. However, due to the nonconvexity of the NMF objective function with respect to both $\mathbf{X}$ and $\mathbf{W}$, it is difficult to obtain a global optimal solution [5]. In addition, NMF lacks a unique solution. In order to restrict the feasible solution set, various constraints have been incorporated into the NMF framework [5], [6], [8], [12], [18].

The commonly used constraints are sparsity constraints on the abundance matrix $\mathbf{W}$, such as ℓ_q-norm ($0 < q \leq 1$)

constraint on $\mathbf{W}$. The ℓ_q-norm-based NMF unmixing model (ℓ_q-NMF) only changes the objective function of NMF model (3) as

$$\min_{\mathbf{X},\mathbf{W}} \left\|\mathbf{Y}-\mathbf{X}\mathbf{W}\right\|_F^2 + \lambda\|\mathbf{W}\|_q \tag{4}$$

where λ is a regularization parameter and $\|\mathbf{W}\|_q$ is the ℓ_q-regularizer, that is

$$\|\mathbf{W}\|_q = \sum_{n,p=1}^{N,P} (\mathbf{w}_n(p))^q.$$

The ℓ_q-NMF model (4) can be easily solved by the following multiplicative iterative rules:

$$\mathbf{X} = \mathbf{X}.*(\mathbf{Y}\mathbf{W}^{\mathrm{T}})./(\mathbf{X}\mathbf{W}\mathbf{W}^{\mathrm{T}}) \tag{5}$$

$$\mathbf{W} = \mathbf{W}.*(\bar{\mathbf{X}}^{\mathrm{T}}\bar{\mathbf{Y}})./(\bar{\mathbf{X}}^{\mathrm{T}}\bar{\mathbf{X}}\mathbf{W} + \lambda q\mathbf{W}^{q-1}) \tag{6}$$

where $\bar{\mathbf{Y}} = [\mathbf{Y}; \delta\mathbf{1}_N^{\mathrm{T}}]$ and $\bar{\mathbf{X}} = [\mathbf{X}; \delta\mathbf{1}_P^{\mathrm{T}}]$ are the augmented matrices by adding a row of constants to the hyperspectral data matrix $\mathbf{Y}$ and endmember matrix $\mathbf{X}$ for the purpose of imposing the abundance sum-to-one constraint with parameter δ [5].

III. Self-Paced NMF Unmixing Model

A. Self-Paced Learning

Given a set of training samples $\{\mathbf{z}_i, y_i\}_{i=1}^N$, the task of classical learning algorithm is to learn a function f associated with parameter $\boldsymbol{\theta}$ by minimizing the following empirical risk:

$$\min_{\boldsymbol{\theta}} \ \mathbb{E}(\boldsymbol{\theta}) = \sum_{i=1}^{N} L(y_i, f(\mathbf{z}_i, \boldsymbol{\theta})) \tag{7}$$

where L is a loss function which measures the error between the true and estimated labels.

The traditional learning scheme (7) needs to use all training samples to learn model parameter $\boldsymbol{\theta}$. However, in reality, not all samples are useful especially when there exist outliers in the data. In addition, the importance of training samples is also different. Therefore, it would be better to consider the differences in training samples in the design of the learning model. SPL considers the differences in training samples by adaptively selecting samples from easy to complex into training using a self-paced regularizer [21], [22]. In detail, SPL simultaneously optimizes the model parameter $\boldsymbol{\theta}$ and the sample weight vector $\mathbf{u} = [u_1, \ldots, u_N]^{\mathrm{T}}$ as

$$\min_{\boldsymbol{\theta},\mathbf{u}} \ \mathbb{E}(\boldsymbol{\theta}, \mathbf{u}) = \left\{ \sum_{i=1}^{N} \omega_i L\big(y_i, f(\mathbf{z}_i, \boldsymbol{\theta})\big) + h(\gamma, u_i) \right\} \tag{8}$$

where u_i is a weight that describes the degree of complexity of sample $\mathbf{z}_i$, $h(\gamma, \mathbf{u})$ is called as self-paced function, and γ is a "model age" parameter that decides the size of the model [22]. Given a small γ, only "easy" samples with small losses $\ell_i = L\big(y_i, f(\mathbf{z}_i, \boldsymbol{\theta})\big)$ are considered. By gradually increasing the model age γ, more and more complex samples with larger losses are included into the training process to learn a more "mature" model [22]. The solution of SPL model (8) can be obtained by alternatively updating model parameter $\boldsymbol{\theta}$ and weight vector $\mathbf{u}$.

B. Self-Paced NMF for Band Weighting

From the ℓ_q-NMF model (4), we can see that the loss function $\|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2$ is a least-squares metric that is sensitive to noise or outliers in the data $\mathbf{Y}$. Due to the imaging mechanism, the hyperspectral data $\mathbf{Y}$ inevitably exists some noisy or outlying bands. The images corresponding to different spectral bands may be corrupted by different kinds of noise, such as the Gaussian noise, stripe noise, impulse noise, deadlines, and their mixtures. The noisy bands will dramatically affect the stableness and accuracy of the least-squares-based NMF unmixing model. In order to eliminate the negative effect of noisy bands, we introduce SPL strategy to automatically select some important spectral bands for unmixing. For this purpose, we first expand the least-squares loss function in row

$$L = \|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2 = \sum_{b=1}^{B} \left\|\mathbf{Y}^b - (\mathbf{X}\mathbf{W})^b\right\|_2^2 \tag{9}$$

where $\mathbf{Y}^b$ denotes the bth row of the matrix $\mathbf{Y}$.

Then, we embed the SPL strategy (8) into the ℓ_q-NMF model (4) and propose a self-paced NMF unmixing model for band weighting (SpNMFB) as follows:

$$\min_{\mathbf{X},\mathbf{W},\mathbf{u}} \sum_{b=1}^{B} \left\{ u_b \left\|\mathbf{Y}^b - (\mathbf{X}\mathbf{W})^b\right\|_2^2 + h(\gamma, u_b) \right\} + \lambda\|\mathbf{W}\|_q \tag{10}$$

where $\mathbf{u} = [u_1; u_2; \ldots; u_B]$ is a weight vector to describe the importance of different spectral bands, and h is a self-paced function associated with "model age" parameter γ.

In model (10), the self-paced regularizer can be used to adaptively select some "good" bands based on the reconstruction residual of each band. The solution of model (10) can be obtained by an alternative updating strategy.

Fixing the weight vector $\mathbf{u}$, the SpNMFB model can be rewritten as

$$\min_{\widetilde{\mathbf{X}},\mathbf{W}} \left\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{W}\right\|_F^2 + \lambda\|\mathbf{W}\|_q \tag{11}$$

where $\widetilde{\mathbf{Y}} = \mathbf{U}^{(1/2)}\mathbf{Y}$, $\widetilde{\mathbf{X}} = \mathbf{U}^{(1/2)}\mathbf{X}$, and $\mathbf{U}$ is a diagonal matrix whose diagonal element is $\mathbf{U}_{bb} = u_b$. It is clear that model (11) is also an ℓ_q-NMF model and can be solved by the multiplicative update rule. The final endmember matrix is $\mathbf{X} = \mathbf{U}^{-(1/2)}\widetilde{\mathbf{X}}$.

When $\mathbf{X}$ and $\mathbf{W}$ are obtained, the weight u_b for the bth spectral band can be optimized by

$$\min_{0\le u_b\le 1} u_b\ell_b + h(\gamma, u_b) \tag{12}$$

where $\ell_b = \|\mathbf{Y}^b - (\mathbf{X}\mathbf{W})^b\|_2^2$ is a constant for given $\mathbf{X}$ and $\mathbf{W}$. Thus, the weight is only determined by the self-paced function h. The commonly used self-paced functions are binary, linear, logarithmic, and mixture functions [17], [22]. Taking the following binary function as an example

$$h(\gamma, u_b) = -\gamma u_b \tag{13}$$

the weight vector of model (12) is

$$u_b = \begin{cases} 1, & \ell_b \le \gamma \\ 0, & \ell_b > \gamma. \end{cases} \tag{14}$$

It is clear that the spectral bands with small errors ($\ell_b \leq \gamma$) are used, and bands whose losses are larger than the threshold γ are not used for unmixing ($u_b = 0$). The weight vector $\mathbf{u}$ indicates the importance or complexity of different spectral bands. Initializing the model age γ as a small value, only very limited "easy" or "good" bands are involved in the hyperspectral unmixing model. By gradually increasing γ, more bands from easy to the complex are included in the learning process [22].

In the implementation of the algorithm, given the initial endmember matrix $\mathbf{X}_0$ and the abundance matrix $\mathbf{W}_0$, SpNMFB iteratively updates the weight vector, endmember, and abundance matrices and finally obtains the endmember matrix $\mathbf{X}^*$ and the abundance matrix $\mathbf{W}^*$. To refine the results, we set the obtained $\mathbf{X}^*$ and $\mathbf{W}^*$ as the initial endmember and abundance matrices to rerun the algorithm. By repeating the process several times, the performance of our SpNMFB algorithm can be improved to a certain extent. The detailed process of the proposed SpNMFB algorithm can be summarized in Algorithm 1.

Algorithm 1 Self-Paced NMF Model for Band Weighting (SpNMFB)

Input: Hyperspectral data matrix $\mathbf{Y}$,
initial endmember $\mathbf{X}_0$ and abundance $\mathbf{W}_0$,
initial model age γ_0 and step size η,
the number of repetitions R.
Output: Endmember and abundance matrices.

Repeatedly run the following steps R times:
1. Initialize $\mathbf{X}^{(0)} = \mathbf{X}_0$, $\mathbf{W}^{(0)} = \mathbf{W}_0$, $\gamma^{(0)} = \gamma_0$, set $k = 1$.
2. Run the following steps until convergence:
   (a) calculate the band reconstruction errors:
$$\ell_b^{(k)} = \|\mathbf{Y}^b - (\mathbf{X}^{(k-1)}\mathbf{W}^{(k-1)})^b\|_2^2$$
   (b) estimate self-paced weights:
$$u_b^{(k)} = \arg\min_{0\leq u_b\leq 1} u_b\ell_b^{(k)} + h(\gamma^{(k-1)}, u_b)$$
   (c) update weight: $\mathbf{U}^{(k)} = \mathrm{diag}\{u_1^{(k)}, \ldots, u_B^{(k)}\}$
   (d) compute matrix:
$$\widetilde{\mathbf{Y}} = (\mathbf{U}^{(k-1)})^{\frac{1}{2}}\mathbf{Y}$$
$$\widetilde{\mathbf{X}}^{(k-1)} = (\mathbf{U}^{(k-1)})^{\frac{1}{2}}\mathbf{X}^{(k-1)}$$
   (e) update endmember and abundance:
$$(\widetilde{\mathbf{X}}^{(k)}, \mathbf{W}^{(k)}) = \mathrm{L_qNMF}(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}^{(k-1)}, \mathbf{W}^{(k-1)})$$
$$\mathbf{X}^{(k)} = (\mathbf{U}^{(k-1)})^{-\frac{1}{2}}\widetilde{\mathbf{X}}^{(k)}$$
   (f) update model age: $\gamma^{(k)} = \gamma^{(k-1)} + \eta$
   (g) $k \leftarrow k + 1$
3. Set $\mathbf{X}_0 = \mathbf{X}^{(k-1)}$, $\mathbf{W}_0 = \mathbf{W}^{(k-1)}$

C. Self-Paced NMF for Pixel Weighting

Because ℓ_q-NMF model (4) needs to handle the whole hyperspectral data matrix $\mathbf{Y}$, the noisy or background pixels in $\mathbf{Y}$ also affect the unmixing performance. To eliminate the effect of noisy pixels, we propose a self-paced NMF unmixing model for pixel weight (SpNMFP) as follows:

$$\min_{\mathbf{X},\mathbf{W},\mathbf{u}} \sum_{n=1}^{N}\{u_n\|\mathbf{Y}_n - (\mathbf{XW})_n\|_2^2 + h(\gamma, u_n)\} + \lambda\|\mathbf{W}\|_q \quad (15)$$

where $\mathbf{Y}_n$ denotes the nth column of the matrix $\mathbf{Y}$, $\mathbf{u} = [u_1; u_2; \ldots; u_N]$ is a weight vector to describe the importance of different pixels, and h is a self-paced function associated with "model age" parameter γ.

The solution to model (15) can also be obtained by an alternative updating strategy. Fixeing the weight vector $\mathbf{u}$, the SpNMFP model can be rewritten as

$$\min_{\mathbf{X},\widetilde{\mathbf{W}}} \left\|\widetilde{\mathbf{Y}} - \mathbf{X}\widetilde{\mathbf{W}}\right\|_F^2 + \lambda\|\mathbf{W}\|_q \quad (16)$$

where $\widetilde{\mathbf{Y}} = \mathbf{YU}^{\frac{1}{2}}$, $\widetilde{\mathbf{W}} = \mathbf{WU}^{(1/2)}$, and $\mathbf{U}$ is a diagonal matrix whose diagonal element is $\mathbf{U}_{nn} = u_n$. As $\mathbf{W}$ and $\widetilde{\mathbf{W}}$ have similar sparsity structure and $\|\widetilde{\mathbf{W}}\|_q \leq \|\mathbf{W}\|_q$, the model (16) can be modified as

$$\min_{\mathbf{X},\widetilde{\mathbf{W}}} \left\|\widetilde{\mathbf{Y}} - \mathbf{X}\widetilde{\mathbf{W}}\right\|_F^2 + \lambda\|\widetilde{\mathbf{W}}\|_q. \quad (17)$$

It is clear that model (17) is also an ℓ_q-NMF unmixing model and can be solved by the multiplicative update rule. The final abundance matrix is $\mathbf{W} = \widetilde{\mathbf{W}}\mathbf{U}^{-\frac{1}{2}}$.

When $\mathbf{X}$ and $\mathbf{W}$ are obtained, the weight u_n for the pixel n can be optimized by

$$\min_{0\leq u_n\leq 1} u_n\ell_n + h(\gamma, u_n) \quad (18)$$

where $\ell_n = \|\mathbf{Y}_n - (\mathbf{XW})_n\|_2^2$ is a constant for given $\mathbf{X}$ and $\mathbf{W}$.

The detailed process of the proposed SpNMFP algorithm can be summarized in Algorithm 2.

D. Self-Paced NMF for Element Weighting (SpNMFE)

The SpNMFB and SpNMFP algorithms deal with the band and pixel noise in the row and column manners, respectively. SpNMFB assigns a weight for each row (i.e., a band), while SpNMFP gives each column (i.e., a pixel) a weight. They consider either a row or a column as a processing unit. However, the noise may not exist in an entire row or column of hyperspectral data $\mathbf{Y}$. In reality, there usually exist noises in some specific spectral bands and spatial positions. This means that some elements of matrix $\mathbf{Y}$, but not its whole row or column, are corrupted by noise. In order to deal with the elementwise noise, we propose an Self-paced NMF model for Element weighting (SpNMFE) as follows:

$$\min_{\mathbf{X},\mathbf{W},\mathbf{u}} \sum_{b,n=1}^{B,N}\{u_{bn}[\mathbf{Y}_{bn} - (\mathbf{XW})_{bn}]^2 + h(\gamma, u_{bn})\} + \lambda\|\mathbf{W}\|_q. \quad (19)$$

The model (19) can also be solved by an alternative updating strategy. Fixing the weight u_{bn}, the SpNMFE model can be rewritten as

$$\min_{\mathbf{X},\mathbf{W}} \sum_{b,n=1}^{B,N}\{u_{bn}[\mathbf{Y}_{bn} - (\mathbf{XW})_{bn}]^2\} + \lambda\|\mathbf{W}\|_q \quad (20)$$

Algorithm 2 Self-Paced NMF Model for Pixel Weighting (SpNMFP)

Input: Hyperspectral data matrix $\mathbf{Y}$,
initial endmember $\mathbf{X}_0$ and abundance $\mathbf{W}_0$,
initial model age γ_0 and step size η,
the number of repetitions R.
Output: Endmember and abundance matrices.
Repeatedly run the following steps R times:
1. Initialize $\mathbf{X}^{(0)} = \mathbf{X}_0$, $\mathbf{W}^{(0)} = \mathbf{W}_0$, $\gamma^{(0)} = \gamma_0$, set $k = 1$.
2. Run the following steps until convergence:
(a) calculate the pixel reconstruction errors:

$$\ell_n^{(k)} = \|\mathbf{Y}_n - (\mathbf{X}^{(k-1)}\mathbf{W}^{(k-1)})_n\|_2^2$$

(b) estimate self-paced weights:

$$u_n^{(k)} = \arg\min_{0\le u_n\le 1} u_n \ell_n^{(k)} + h(\gamma^{(k-1)}, u_n)$$

(c) update weight: $\mathbf{U}^{(k)} = \mathrm{diag}\{u_1^{(k)}, \ldots, u_N^{(k)}\}$
(d) compute matrix:

$$\widetilde{\mathbf{Y}} = \mathbf{Y}(\mathbf{U}^{(k-1)})^{\frac{1}{2}}$$
$$\widetilde{\mathbf{W}}^{(k-1)} = \mathbf{W}^{(k-1)}(\mathbf{U}^{(k-1)})^{\frac{1}{2}}$$

(e) update endmember and abundance:

$$(\mathbf{X}^{(k)}, \widetilde{\mathbf{W}}^{(k)}) = \mathrm{L_qNMF}(\widetilde{\mathbf{Y}}, \mathbf{X}^{(k-1)}, \widetilde{\mathbf{W}}^{(k-1)})$$
$$\mathbf{W}^{(k)} = \widetilde{\mathbf{W}}^{(k)}(\mathbf{U}^{(k-1)})^{-\frac{1}{2}}$$

(f) update model age: $\gamma^{(k)} = \gamma^{(k-1)} + \eta$
(g) $k \leftarrow k + 1$
3. Set $\mathbf{X}_0 = \mathbf{X}^{(k-1)}$, $\mathbf{W}_0 = \mathbf{W}^{(k-1)}$

which is a weighted ℓ_q-NMF model (ℓ_q-WNMF) [23]. Imposing additional sum-to-one constraint to the abundance matrix $\mathbf{W}$, the solution to (20) can be obtained by alternatively updating the following two terms

$$\mathbf{X} = \mathbf{X}. * \frac{(\mathbf{U}. * \mathbf{Y})\mathbf{W}^{\mathrm{T}}}{(\mathbf{U}. * (\mathbf{X}\mathbf{W}))\mathbf{W}^{\mathrm{T}}} \tag{21}$$

$$\mathbf{W} = \mathbf{W}. * \frac{\bar{\mathbf{X}}^{\mathrm{T}}(\bar{\mathbf{U}}. * \bar{\mathbf{Y}})}{\bar{\mathbf{X}}^{\mathrm{T}}(\bar{\mathbf{U}}. * (\bar{\mathbf{X}}\mathbf{W})) + \lambda q \mathbf{W}^{q-1}} \tag{22}$$

where $\mathbf{U} = \{u_{bn}\}_{b,n=1}^{B,N}$ is a weight matrix, and $\bar{\mathbf{U}} = [\mathbf{U}; \delta \mathbf{1}_N^{\mathrm{T}}]$.

When $\mathbf{X}$ and $\mathbf{W}$ are obtained, the weight u_{bn} can be optimized by

$$\min_{0\le u_{bn}\le 1} u_{bn}\ell_{bn} + h(\gamma, u_{bn}) \tag{23}$$

where $\ell_{bn} = (\mathbf{Y}_{bn} - (\mathbf{X}\mathbf{W})_{bn})^2$ is a constant for given $\mathbf{X}$ and $\mathbf{W}$.

The detailed process of the proposed SpNMFE algorithm can be summarized in Algorithm 3.

Algorithm 3 Self-Paced NMF Model for Element Weighting (SpNMFE)

Input: Hyperspectral data matrix $\mathbf{Y}$,
initial endmember $\mathbf{X}_0$ and abundance $\mathbf{W}_0$,
initial model age γ_0 and step size η.
Output: Endmember and abundance matrices.
Initialize $\mathbf{X}^{(0)} = \mathbf{X}_0$, $\mathbf{W}^{(0)} = \mathbf{W}_0$, $\gamma^{(0)} = \gamma_0$, set $k = 1$.
Run the following steps until convergence:
(a) calculate elementwise errors:

$$\ell_{bn}^{(k)} = \left(\mathbf{Y}_{bn} - (\mathbf{X}^{(k-1)}\mathbf{W}^{(k-1)})_{bn}\right)^2$$

(b) estimate self-paced weights:

$$u_{bn}^{(k)} = \arg\min_{0\le u_{bn}\le 1} u_{bn}\ell_{bn}^{(k)} + h(\gamma^{(k-1)}, u_{bn})$$

(c) update weight: $\mathbf{U}^{(k)} = \{u_{bn}^{(k)}\}_{b,n=1}^{B,N}$
(d) update endmember and abundance:

$$(\mathbf{X}^{(k)}, \mathbf{W}^{(k)}) = \mathrm{L_qWNMF}(\mathbf{Y}, \mathbf{X}^{(k-1)}, \mathbf{W}^{(k-1)}, \mathbf{U}^{(k)})$$

(e) update model age: $\gamma^{(k)} = \gamma^{(k-1)} \times \eta$
(f) $k \leftarrow k + 1$

IV. Experiments

In this section, our proposed ℓ_q-norm-based SpNMF methods ($q = 1/2$ for simplicity) are tested on both synthetic and real-world data sets and compared with the following unmixing methods: NMF [19], $\ell_{1/2}$-NMF [5], $\ell_{2,1}$-NMF [24], correntropy-based NMF (CENMF) [12], correntropy-induced metric-based NMF (CIM-NMF) [25], and Huber NMF (Huber-NMF) [25]. For all the NMF-based methods, the VCA and fully constrained least-squares (FCLS) methods are used to initialize the endmember matrix $\mathbf{X}_0$ and abundance matrix $\mathbf{W}_0$, respectively, and the parameter δ for the abundance sum-to-one constraint is set as 15 [6]. The regularization parameter λ in $\ell_{1/2}$-NMF or CENMF is dependent on the sparsity of the material abundances and is estimated based on the sparseness criterion [5]. In the CIM-NMF, the kernel size is computed as an average reconstruction error [25]. In Huber-NMF, the cutoff parameter in the Huber loss function is set as the median of reconstruction errors [25]. The spectral angle distance (SAD) and the root-mean-square error (RMSE) are used to evaluate the performance of different unmixing methods. The SAD measures the distance between the true and estimated endmember as follows:

$$\mathrm{SAD}_k = \arccos\left(\frac{\mathbf{x}_k^{\mathrm{T}}\widetilde{\mathbf{x}}_k}{\|\mathbf{x}_k\|\|\widetilde{\mathbf{x}}_k\|}\right) \tag{24}$$

where $\mathbf{x}_k$ and $\widetilde{\mathbf{x}}_k$ are the kth reference and the estimated endmember signatures, respectively. The RMSE is used to measure the distance between the ground-truth abundance map and its estimated abundance map as follows:

$$\mathrm{RMSE}_k = \left(\frac{1}{N}\left\|\mathbf{w}_k - \widetilde{\mathbf{w}}_k\right\|_2^2\right)^{\frac{1}{2}} \tag{25}$$

where $\mathbf{w}_k$ and $\widetilde{\mathbf{w}}_k$ are the kth true and estimated abundance maps in the vector form.

As recommended in [17], the vectorwise SpNMF methods (i.e., SpNMFB and SpNMFP) use a mixture self-paced function [22]

$$h(\gamma, u_t) = -\zeta \log\left(u_t + \frac{\zeta}{\gamma_1}\right), \quad \zeta = \frac{\gamma_1\gamma_2}{\gamma_1 - \gamma_2},\ 0 < \gamma_2 < \gamma_1 \tag{26}$$

and the corresponding weight has the form

$$u_t = f(\ell_t) = \begin{cases} 1, & \ell_t \le \gamma_2 \\ 0, & \ell_t \ge \gamma_1 \\ \zeta(\gamma_1 - \ell_t)/(\gamma_1 \ell_t), & \gamma_2 < \ell_t < \gamma_1 \end{cases} \tag{27}$$

where γ_1 and γ_2 are model age parameters.

It can be seen from (27) that the model age parameters are used to bound the errors such that the bands or pixels with smaller errors are assigned larger weights and vice visa. As the range of error $\boldsymbol{\ell}$ is data-dependent, it is difficult to determine its upper and lower bounds γ_1 and γ_2 directly. Considering that the model age parameters are used to define the current model size, we can directly determine the number of bands or pixels involved in each learning process according to their errors [17]. For this purpose, we predefine an increasing sequence $\{T_1, T_2, \ldots, T_{\max}\}$ with $T_i = \lfloor k_1^i T \rfloor$ and $T_{\max} = T$ ($T = B$ for SpNMFB and $T = N$ for SpNMFP), where T_i denotes the number of bands or pixels selected in the ith SPL process. In the ith iteration, we sort the loss vector $\boldsymbol{\ell}^{(i)}$ in an ascending order, obtain the sorted loss $\boldsymbol{\ell}_a^{(i)}$, and then set the current model age as $\gamma_1^{(i)} = \boldsymbol{\ell}_a^{(i)}(T_i)$ and $\gamma_2^{(i)} = \boldsymbol{\ell}_a^{(i)}(\lfloor k_2 T \rfloor)$. That is, in the ith iteration, there are $\lfloor k_2 T \rfloor$ absolutely easy (weight 1) and $T_{\max} - T_i$ absolutely complex (weight 0) bands or pixels. In the experiment, we define $k_1^i = k_1 + (i-1)\delta$, where k_1 corresponds to the initial model age and is set as $k_1 = 0.5$, and the step size $\delta = 0.05$. Here, k_2 decides the number of atoms to be deleted and is fixed as 0.2. The number of repetitions R is set as 10.

By the integration of SPL weight function $u_t = f(\ell_t)$, the latent SPL mixture loss function can be obtained as [16], [17]

$$F^M(\ell) = \begin{cases} \ell, & \ell \le \gamma_2 \\ \gamma_2 + \zeta\left(\log\dfrac{\ell}{\gamma_2} + \dfrac{\gamma_2}{\gamma_1} - \dfrac{\ell}{\gamma_1}\right), & \gamma_2 < \ell < \gamma_1 \\ \gamma_2 + \zeta\left(\log\dfrac{\gamma_1}{\gamma_2} + \dfrac{\gamma_2}{\gamma_1} - 1\right), & \ell \ge \gamma_1. \end{cases} \tag{28}$$

The elementwise SpNMF method (i.e., SpNMFE) uses another mixture self-paced function [16], [26], [27]

$$\widetilde{h}(\gamma, u_{bn}) = \frac{1}{u_{bn} + 1/\gamma} \tag{29}$$

and the corresponding weight has the form

$$u_{bn} = \widetilde{f}(\ell_{bn}) = \begin{cases} 1, & \ell_{bn} \le \left(\dfrac{\gamma}{\gamma+1}\right)^2 \\ 0, & \ell_{bn} \ge \gamma^2 \\ \left(\dfrac{1}{\sqrt{\ell_{bn}}} - \dfrac{1}{\gamma}\right), & \text{otherwise} \end{cases} \tag{30}$$

where the model age parameter γ is set as $\gamma = \gamma \times \eta$ with $\eta = 1.05$ [26], and the initial value of γ is set as $\gamma_0 = (\text{mean}\{\ell_{bn}, b = 1, \ldots, B, n = 1, \ldots, N\})^{1/2}$.

A. Comparison of Different NMF Unmixing Methods

Denote error matrix as $\mathbf{E} = \mathbf{Y} - \mathbf{XW} \in R^{B \times N}$ and two vectors consisting of the ℓ_2 norm of its row and column vectors as

$$\mathbf{e}_r = (\|\mathbf{E}^1\|_2, \|\mathbf{E}^2\|_2, \ldots, \|\mathbf{E}^B\|_2)^{\mathrm{T}} \in R^{B \times 1}$$
$$\mathbf{e}_c = (\|\mathbf{E}_1\|_2, \|\mathbf{E}_2\|_2, \ldots, \|\mathbf{E}_N\|_2)^{\mathrm{T}} \in R^{N \times 1}.$$

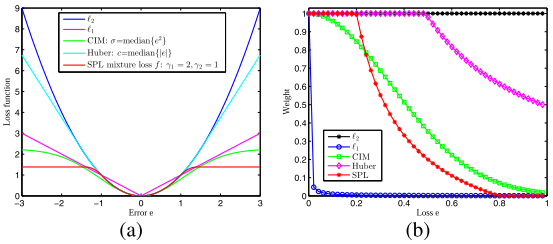


Fig. 1. Comparison of different loss function. (a) Loss function. (b) Weight.

We show the loss function and regularizer of different NMF unmixing methods in Table I. It can be seen that ℓ_q-NMF improves the original NMF by adding ℓ_q-regularizer on the abundance matrix $\mathbf{W}$, while other NMF methods modify the least-squares-based F-norm loss function of NMF to robust loss functions, such as $\ell_{2,1}$-norm loss, correntropy-based loss, Huber loss, and SPL losses. These robust NMF unmixing methods can be transferred to weighted NMF methods, and their corresponding weights are shown in the last column of Table I. They have different weighting strategies: 1) $\ell_{2,1}$-NMF and SpNMFP can be considered as columnwise pixel-weighted methods; 2) CENMF and SpNMFB are rowwise band-weighted methods; and 3) CIM-NMF, Huber-NMF, and SpNMFE are elementwise weighted methods.

To further compare different robust unmixing methods, Fig. 1 shows the corresponding robust loss functions and weights. It is obvious that the robust losses, i.e., Huber, ℓ_1, CIM, and SPL mixture loss (28), can suppress the effect of large errors. In particular, our SPL loss directly truncates the loss of larger errors to a constant such that the effect of large error elements (i.e., noise or outliers) is minimized. From the weights, it can be seen that the ℓ_1 weights are dominated by the errors around 0. The Huber weight is a combination of ℓ_2 and ℓ_1 weights and has a suppression effect for large errors. Our SPL weight consists of three parts: ℓ_2 weight (i.e., weight 1) for small errors, ℓ_1 weight for moderate errors, and 0 weight for large errors.

B. Experiments on Synthetic Data

Here, we select seven spectral signatures from the United States Geological Survey (USGS) digital spectral library as endmembers to generate the synthetic data. The spectral signatures with reflectance values in 224 spectral bands in the range 0.4–2.5 μm are shown in Fig. 2. These seven spectral signatures consist of the endmember matrix $\mathbf{X}$. The abundances are generated based on [6] and [8] and recorded in matrix $\mathbf{W}$. By multiplying the endmember matrix $\mathbf{X}$ and the abundance matrix $\mathbf{W}$, we obtain the synthetic data $\mathbf{Y} = \mathbf{XW}$.

In order to evaluate the robustness of different algorithms, the Gaussian noise is added to the synthetic data $\mathbf{Y}$. The following three experiments are conducted to investigate the effect of different kinds of noise.

TABLE I
COMPARISON OF DIFFERENT NMF UNMIXING METHODS

NMF methods	Loss function	Regularizer	Weight
NMF [19]	$\ell(\mathbf{E}) = \|\mathbf{E}\|_F^2 = \sum_{b=1}^{B}\sum_{n=1}^{N} E_{bn}^2$		$u_{bn} = 1$
ℓ_q-NMF [5]	$\ell(\mathbf{E}) = \|\mathbf{E}\|_F^2 = \sum_{b=1}^{B}\sum_{n=1}^{N} E_{bn}^2$	$\|\mathbf{W}\|_q$	$u_{bn} = 1$
$\ell_{2,1}$-NMF [4], [24]	$\ell(\mathbf{E}) = \|\mathbf{E}\|_{2,1} = \|\mathbf{e}_c\|_1 = \sum_{n=1}^{N}\sqrt{\sum_{b=1}^{B} E_{bn}^2}$		$u_n = \frac{1}{\mathbf{e}_{c,n}}, n = 1, \cdots, N$
CENMF [12]	$\ell(\mathbf{e}_r) = \sum_{b=1}^{B} g(\mathbf{e}_{r,b}), g(e) = 1 - \exp(-\frac{e^2}{\sigma^2})$	$\|\mathbf{W}\|_1$	$u_b = \exp(-\frac{\mathbf{e}_{r,b}^2}{\sigma^2}), b = 1, \cdots, B$
CIM-NMF [25]	$\ell(\mathbf{E}) = \sum_{b=1}^{B}\sum_{n=1}^{N} g(E_{bn}), g(e) = 1 - \exp(-\frac{e^2}{\sigma^2})$		$u_{bn} = \exp(-\frac{E_{bn}^2}{\sigma^2})$
Huber-NMF [25]	$\ell(\mathbf{E}) = \sum_{b=1}^{B}\sum_{n=1}^{N} p(E_{bn}), p(e) = \begin{cases} e^2, & \|e\| \le c \\ 2c\|e\| - c^2, & \|e\| > c \end{cases}$		$u_{bn} = \begin{cases} 1, & \|E_{bn}\| \le c \\ c/\|E_{bn}\|, & \|E_{bn}\| > c \end{cases}$
SpNMFB	$\ell(\mathbf{e}_r) = \sum_{b=1}^{B}\left\{u_b \mathbf{e}_{r,b}^2 + h(\gamma, u_b)\right\}$	$\|\mathbf{W}\|_q$	$u_b = f(\mathbf{e}_{r,b}^2), b = 1, \cdots, B$
SpNMFP	$\ell(\mathbf{e}_c) = \sum_{n=1}^{N}\left\{u_n \mathbf{e}_{c,n}^2 + h(\gamma, u_n)\right\}$	$\|\mathbf{W}\|_q$	$u_n = f(\mathbf{e}_{c,n}^2), n = 1, \cdots, N$
SpNMFE	$\ell(\mathbf{E}) = \sum_{b=1}^{B}\sum_{n=1}^{N}\left\{u_{bn} E_{bn}^2 + \widetilde{h}(\gamma, u_{bn})\right\}$	$\|\mathbf{W}\|_q$	$u_{bn} = \widetilde{f}(E_{bn}^2)$

TABLE II
SAD AND THE STANDARD DERIVATION (IN PERCENT) RESULTS IN THE CASE OF BAND NOISE

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
10	0.3563±5.50	0.3210±5.66	0.3050±4.93	0.2985±5.25	0.2920±4.25	0.2988±4.77	0.3047±6.13	**0.2685±4.43**	0.3687±5.44
15	0.2058±4.70	0.1812±4.37	0.1789±4.04	0.1696±3.88	0.1822±3.65	0.1763±3.89	0.1591±5.08	**0.1560±4.08**	0.1876±6.83
20	0.0819±4.23	0.0766±3.83	0.0797±3.72	0.0741±3.88	0.0878±3.53	0.0815±3.63	**0.0558±3.50**	0.0602±3.90	0.0916±4.09
25	0.0187±0.46	0.0183±0.46	0.0189±0.53	0.0186±0.45	0.0218±0.87	0.0197±0.66	0.0139±0.23	**0.0114±0.04**	0.0222±1.25
30	0.0134±0.24	0.0127±0.25	0.0138±0.28	0.0137±0.25	0.0147±0.33	0.0139±0.29	0.0105±0.08	**0.0097±0.02**	0.0227±0.05

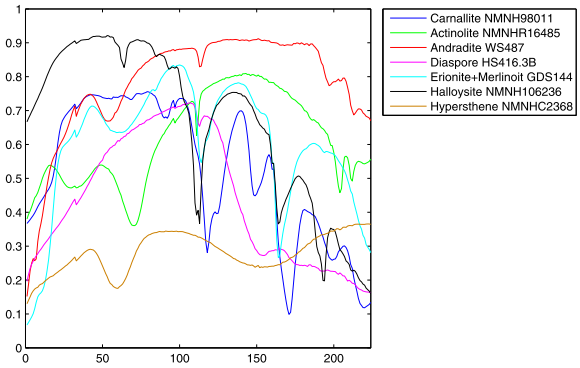


Fig. 2. Spectral signatures of seven endmembers chosen from the USGS library.

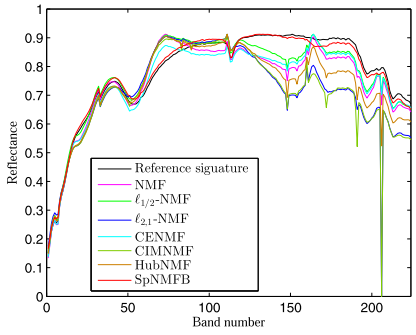


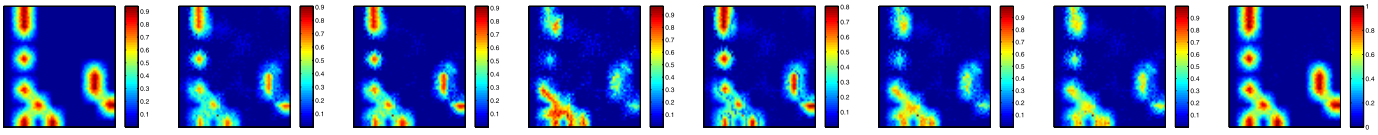
Fig. 3. Reference and estimated spectra for endmember 3 in the case of band noise.

Experiment 1 *(Effect of Band Noise):* In this experiment, the Gaussian noise is added to all spectral bands. The SNRs of bands are generated using the normal distribution SNR $\sim \mathcal{N}(\overline{\text{SNR}}, \epsilon^2)$, where $\overline{\text{SNR}} \in \{10, 15, 20, 25, 30\}$ and $\epsilon = 5$.

Mean and standard derivation of SAD results over 20 randomly runs are reported in Table II. It can be clearly seen that the performance of each algorithm improves as the increase of SNR. Robust unmixing algorithms, i.e., $\ell_{2,1}$-NMF, CENMF, CIM-NMF, and Huber-NMF, improve the unmixing performance of NMF in the case of large noise (i.e., small SNR). However, in the case of small noise, these algorithms generate similar results with the original NMF and cannot suppress the effect of noise. Our proposed SpNMF algorithms (i.e., SpNMFE and SpNMFB) provide better results than other methods. Especially, our SpNMFB produces the best results in different amplitudes of band noise.

Compared with the original NMF, $\ell_{1/2}$-NMF, $\ell_{2,1}$-NMF, CENMF, CIM-NMF, and Huber-NMF improve the SAD results in Table II. However, they do not improve the RMSE results, as shown in Table III. In contrast, our proposed SpNMFB shows excellent performance in terms of both SAD and RMSE.

To visualize the results, the reference and estimated spectra for endmember 3 (i.e., "Andradite WS487") are shown in Fig. 3. It can be clearly seen that the comparison methods

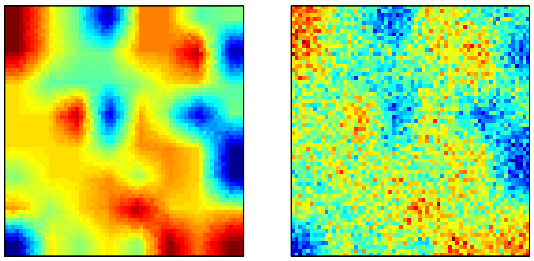


(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4. Abundance maps for endmember 3 in the case of band noise. (a) Ground-truth. (b) NMF. (c) $\ell_{1/2}$-NMF. (d) $\ell_{2,1}$-NMF. (e) CENMF. (f) CIMNMF. (g) HubNMF. (h) SpNMFB.

TABLE III
RMSE AND THE STANDARD DERIVATION (IN PERCENT) RESULTS IN THE CASE OF BAND NOISE

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
10	0.2475±1.71	0.2530±1.95	0.2509±1.91	0.2419±2.18	0.2517±1.90	0.2508±1.93	**0.2418±1.84**	0.2491±2.56	0.2587±2.09
15	0.1958±2.50	0.1973±2.44	0.1975±2.32	0.1841±2.55	0.2021±2.40	0.1973±2.33	0.1761±2.96	**0.1579±2.88**	0.1920±4.29
20	0.1187±4.26	0.1204±4.38	0.1218±4.36	0.1165±4.31	0.1217±4.36	0.1217±4.36	0.0856±5.24	**0.0721±4.61**	0.1248±4.42
25	0.0382±0.96	0.0382±1.01	0.0386±1.03	0.0381±0.98	0.0386±1.03	0.0386±1.03	0.0251±0.43	**0.0211±0.12**	0.0327±2.72
30	0.0333±0.70	0.0349±0.81	0.0361±0.96	0.0344±0.81	0.0362±0.97	0.0361±0.96	0.0195±0.27	**0.0139±0.07**	0.0243±0.26



(a) (b)

Fig. 5. Image of band 51. (a) Original image. (b) Noisy image.

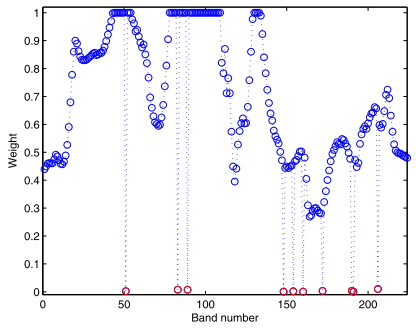


Fig. 6. Weight of different bands learned by the SpNMFB, where the red circles denote noisy bands.

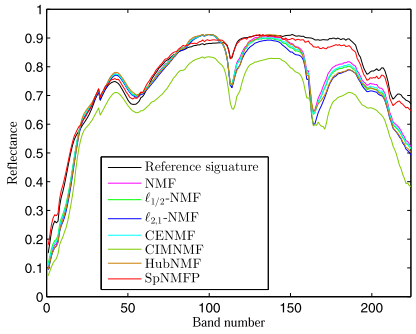


Fig. 7. Reference and estimated spectra for endmember 3 in the case of pixel noise.

produce large derivations after the band number 100, while the estimated spectral curve by our SpNMFB is always consistent with the original reference spectral curve. Fig. 4 shows the abundance map of endmember 3 obtained by different methods, where the map of our SpNMFB is almost the same as the original ground-truth map.

To investigate the band selection ability of SpNMFB, we randomly select ten bands (i.e., 51, 83, 89, 148, 154, 160, 172, 190, 191, and 206) and add Gaussian noise with SNR $\sim \mathcal{N}(15, 5^2)$ to these bands. For a noisy band 51, the corresponding original image and noisy image are shown in Fig. 5, where the information in the original image is almost lost in the noisy image. In this case, the unmixing result will definitely be affected by the noise. Thus, it is better to eliminate or alleviate the effect of these noisy bands. Our proposed SpNMFB method uses an SPL strategy to assign a weight for each band. The weights can reflect the quality of different bands, as shown in Fig. 6. The ten noisy bands are assigned a small weight value 0. This demonstrates that our SpNMFB can effectively select some "good" or "easy" bands into the unmixing model and is robust to band noise.

Experiment 2 *(Effect of Pixel Noise):* In this experiment, the Gaussian noise is added to all pixels. The SNRs of different pixels are generated using the normal distribution SNR $\sim \mathcal{N}(\overline{\text{SNR}}, \epsilon^2)$, where $\overline{\text{SNR}} \in \{10, 15, 20, 25, 30\}$ and $\epsilon = 5$. The mean and standard derivation of SAD and RMSE results are shown in Tables IV and V, respectively. It can be clearly seen that our proposed SpNMFP algorithm provides the best overall results in the case of pixel noise.

The reference and estimated spectra for endmember 3 are shown in Fig. 7, where the spectral curve of our SpNMFP almost coincides with the original reference spectral curve. Fig. 8 shows the abundance map of endmember 3 obtained by different methods. It is clear that the map of our SpNMFP is almost the same as the original ground-truth map, while other algorithms produce a false yellow region on the right upper side of the map.

TABLE IV

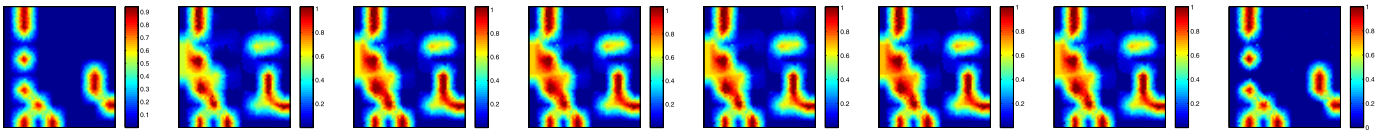
SAD AND THE STANDARD DERIVATION (IN PERCENT) RESULTS IN THE CASE OF PIXEL NOISE

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
10	0.4892±4.97	0.4643±4.47	0.4797±5.15	0.4805±4.84	0.4769±5.17	0.4766±5.11	0.4388±5.09	0.4422±5.35	**0.4335±4.55**
15	0.2395±7.42	0.2217±7.06	0.2306±2.72	0.2318±6.50	0.2273±5.53	0.2306±5.71	**0.2022±8.01**	0.2226±7.94	0.2069±7.70
20	0.0610±2.50	0.0551±2.18	0.0625±2.72	0.0619±2.49	0.0649±2.97	0.0630±2.70	0.0424±2.36	0.0372±2.25	**0.0223±1.60**
25	0.0433±4.38	0.0386±3.51	0.0352±1.94	0.0418±3.62	0.0352±1.74	0.0354±1.91	0.0356±4.35	0.0360±2.26	**0.0312±2.26**
30	0.0186±0.59	0.0178±0.55	0.0199±0.76	0.0207±0.79	0.0223±1.22	0.0203±0.77	0.0152±0.60	0.0126±0.52	**0.0095±0.07**

TABLE V

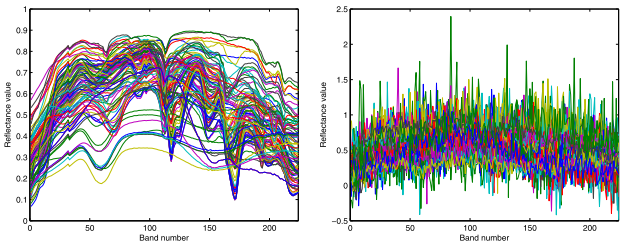
RMSE AND THE STANDARD DERIVATION (IN PERCENT) RESULTS IN THE CASE OF PIXEL NOISE

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
10	0.2467±1.78	0.2496±2.30	0.2482±2.10	0.2466±1.98	0.2484±2.11	0.2485±2.09	**0.2450±2.04**	0.2504±2.55	0.2469±2.01
15	0.1922±3.54	0.1996±3.86	0.2020±3.24	0.1984±3.27	0.2010±3.14	0.2020±3.24	0.1822±4.74	0.2002±4.82	**0.1603±3.48**
20	0.0884±2.25	0.0871±2.24	0.0947±3.07	0.0922±2.42	0.0936±2.68	0.0948±3.07	0.0648±2.46	0.0452±2.03	**0.0431±1.46**
25	0.0655±4.11	0.0642±3.69	0.0638±3.07	0.0663±3.72	0.0639±3.06	0.0638±3.07	0.0461±4.44	0.0377±4.07	**0.0375±2.56**
30	0.0413±1.59	0.0442±1.76	0.0474±2.11	0.0467±1.99	0.0499±2.95	0.0475±2.10	0.0282±1.23	**0.0236±0.86**	0.0239±0.05



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 8. Abundance maps for endmember 3 in the case of pixel noise. (a) Ground truth. (b) NMF. (c) $\ell_{1/2}$-NMF. (d) $\ell_{2,1}$-NMF. (e) CENMF. (f) CIMNMF. (g) HubNMF. (h) SpNMFP.



(a) (b)

Fig. 9. Spectral curves of selected 100 pixels. (a) Original pixels. (b) Noisy pixels.

To investigate the pixel selection ability of SpNMFP, we randomly select 100 pixels and add the Gaussian noise with SNR $\sim \mathcal{N}(15, 5^2)$ to these pixels. The original and noisy spectral curves of these 100 pixels are shown in Fig. 9, where the spectral characteristics in the original spectra are almost lost in the noisy spectra. To eliminate the effect of pixel noise, our proposed SpNMFP method assigns a weight for each pixel, as shown in Fig. 10. It is obvious that 100 noisy pixels are located in the black region. That is, the weight of noisy pixels is 0. By gradually selecting "good" pixels into the unmixing model, SpNMFP can exclude noisy pixels.

Experiment 3 *(Effect of Element Noise):* In this experiment, the whole synthetic data $\mathbf{Y}$ are degraded by the Gaussian noise with distribution SNR $\sim \mathcal{N}(\overline{\text{SNR}}, \epsilon^2)$, where $\overline{\text{SNR}} \in \{10, 15, 20, 25, 30\}$ and $\epsilon = 5$. In this case, we add the noise to the data matrix $\mathbf{Y}$ simultaneously rather than adding noise to each row or column. The mean and standard derivation

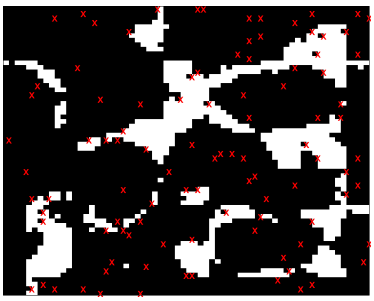


Fig. 10. Weight of different pixels learned by the SpNMFP, where black and white colors denote weights 0 and 1, respectively. The red crosses indicate the position of noisy pixels.

(a) (b)

Fig. 11. Real data sets used in the experiments. (a) Jasper. (b) Urban.

of SAD and RMSE results are shown in Tables VI and VII, respectively. It can be clearly seen that our proposed SpNMFE and SpNMFB provide much better results than other methods in the case of elementwise image noise.

TABLE VI
SAD AND THE STANDARD DERIVATION (IN PERCENT) RESULTS IN THE CASE OF ELEMENT NOISE

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
10	0.0730±1.01	0.0696±0.90	0.0721±0.95	0.0723±0.94	0.0811±2.76	0.0729±0.94	0.0367±0.15	**0.0308±0.17**	0.0467±1.74
15	0.0373±0.62	0.0365±0.62	0.0373±0.66	0.0375±0.65	0.0405±0.68	0.0379±0.66	**0.0203±0.20**	0.0205±0.28	0.0251±0.45
20	0.0216±0.28	0.0212±0.26	0.0218±0.29	0.0219±0.29	0.0234±0.26	0.0221±0.28	0.0146±0.39	**0.0122±0.03**	0.0196±0.22
25	0.0160±0.31	0.0157±0.29	0.0168±0.42	0.0170±0.42	0.0172±0.44	0.0168±0.42	0.0106±0.31	**0.0100±0.18**	0.0218±0.12
30	0.0107±0.38	0.0105±0.34	0.0109±0.45	0.0113±0.46	0.0104±0.41	0.0109±0.46	**0.0076±0.10**	0.0091±0.01	0.0239±0.05

TABLE VII
RMSE AND THE STANDARD DERIVATION (IN PERCENT) RESULTS IN THE CASE OF ELEMENT NOISE

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
10	0.1025±1.05	0.1024±1.17	0.1037±1.21	0.1039±1.21	0.1091±3.10	0.1038±1.22	**0.0704±0.24**	0.0749±0.18	0.0866±1.58
15	0.0667±0.70	0.0667±0.78	0.0673±0.79	0.0674±0.78	0.0674±0.79	0.0673±0.79	**0.0466±0.36**	0.0469±0.22	0.0487±0.36
20	0.0421±0.57	0.0422±0.66	0.0427±0.68	0.0439±0.68	0.0428±0.69	0.0427±0.69	0.0290±0.49	**0.0264±0.04**	0.0305±0.33
25	0.0356±0.61	0.0374±0.74	0.0385±0.92	0.0385±0.90	0.0387±0.92	0.0385±0.92	0.0251±0.60	**0.0164±0.05**	0.0257±0.23
30	0.0277±0.83	0.0296±0.92	0.0302±1.13	0.0301±1.05	0.0304±1.14	0.0303±1.13	0.0191±0.25	**0.0114±0.05**	0.0242±0.15

TABLE VIII
SAD OF DIFFERENT METHODS ON THE JASPER DATA SET WITH THE NOISY BANDS

	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
Tree	0.2781	0.2934	0.3420	0.2219	0.2007	0.2334	0.1484	0.0859	**0.0765**
Water	0.2928	0.1481	0.2837	0.3006	0.2904	0.2763	**0.1085**	0.2144	0.1832
Soil	0.4088	0.1166	0.1163	0.2567	0.2344	0.1371	0.0951	**0.0706**	0.0721
Road	0.5738	0.6275	0.7185	0.7377	0.7332	0.7614	0.5032	0.2523	**0.1810**
Mean	0.3884	0.2964	0.3651	0.3792	0.3647	0.3520	0.2138	0.1558	**0.1282**

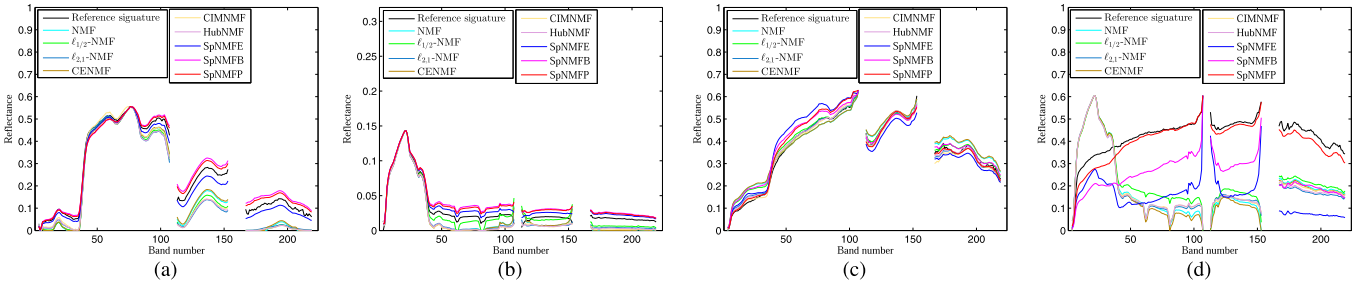


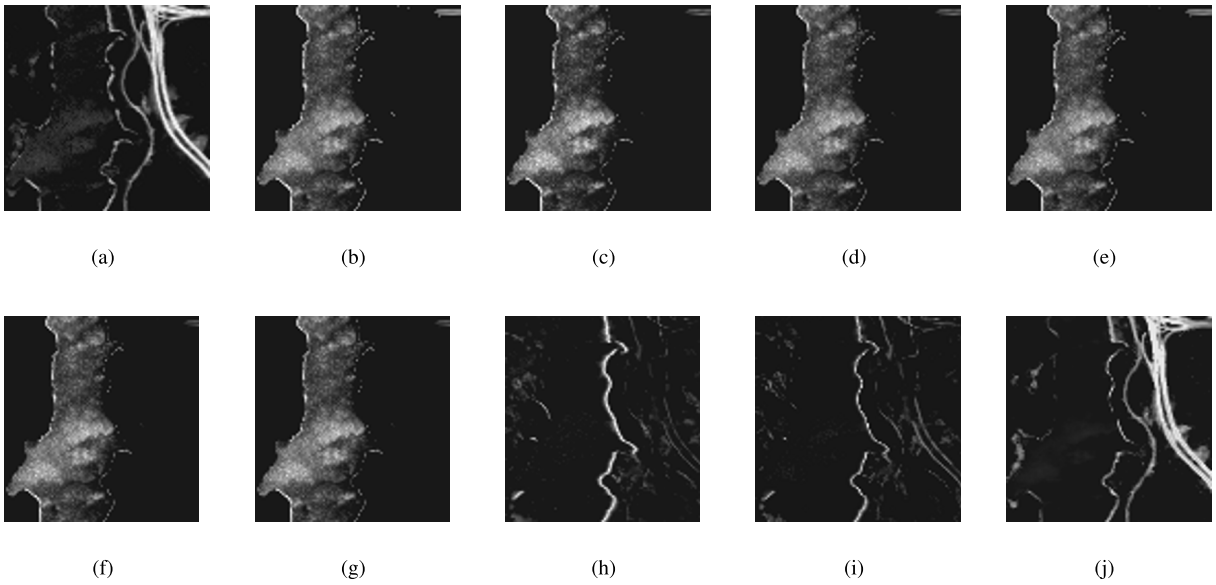
Fig. 12. Endmember signatures estimated by different methods over the Jasper data with the noisy bands. (a) Tree. (b) Soil. (c) Water. (d) Road.

C. Experiments on Real Data

In this section, the performance of different unmixing methods is validated using two real-world hyperspectral data sets, i.e., Jasper and Urban, as shown in Fig. 11.

The first hyperspectral data set is the Jasper data set, which is collected by an airborne visible/infrared imaging spectrometer (AVIRIS) sensor. The data set consists of 224 spectral bands ranging from 380 to 2500 nm. The spectral resolution is up to 9.46 nm. The size of the scene is 512 × 614. In the experiments, a subimage of 100 × 100 pixels is used. There are four targets in the selected region, including road, soil, water, and tree. Due to dense water vapor and atmospheric effects, there exists some noisy bands, including the bands 1–3, 108–112, 154–166, and 220–224. In this experiment, we investigate the performance of algorithms on both data containing noisy band images (i.e., the original data with $B = 224$) and data without noisy band images (i.e., $B = 198$).

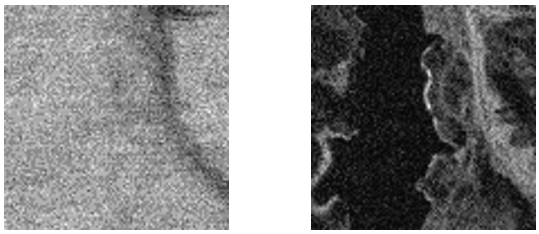
Table VIII shows the SAD results obtained by different methods on the Jasper data set with noisy bands. From the results, we can see that all robust methods improve the original NMF method, and our proposed SpNMFP provides the best result. Although the existing NMF methods slightly improve NMF, they show very poor results on the endmember "Road." In contrast, our SpNMFB and SpNMFP show relatively better results on this endmember. Fig. 12 shows the reference endmember signatures and estimated endmember signatures by different methods over the Jasper data with the noisy bands. For three targets, i.e., Tree, Water, and Soil, the estimated endmember curves by different methods show similar variation tendencies with the reference endmember signa-



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 13. Abundance maps for endmember 4 ("Road") of the Jasper data with the noisy bands. (a) Ground truth, (b) NMF. (c) $\ell_{1/2}$-NMF. (d) $\ell_{2,1}$-NMF. (e) CENMF. (f) CIMNMF. (g) HubNMF. (h) SpNMFE. (i) SpNMFB. (j) SpNMFP.



(a) (b)

Fig. 14. Two noisy bands of Jasper. (a) Band 2. (c) Band 224.

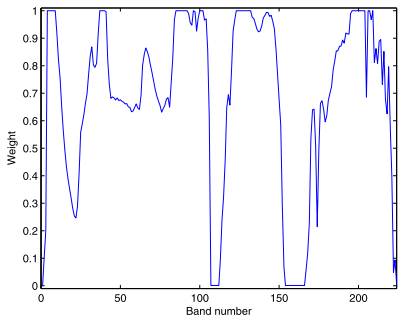


Fig. 15. Band weights of Jasper with the noisy bands.

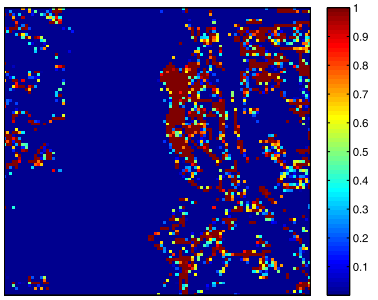
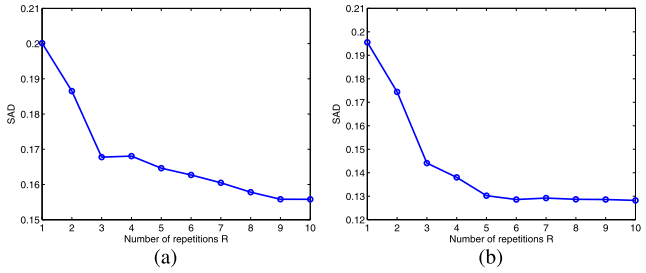


Fig. 16. Pixel weights of Jasper.



(a) (b)

Fig. 17. SAD versus number of repetitions for (a) SpNMFB and (b) SpNMFP on the Jasper data.

tures. Notwithstanding, the results of our proposed SpNMFE, SpNMFB, and SpNMFP methods are obviously more close to the true reference endmembers. On the target "Road," except for SpNMFB and SpNMFP, other methods are almost failed. The abundance maps corresponding to endmember "Road" obtained by different methods are shown in Fig. 13. It can be clearly seen that our proposed SpNMFP provides highly consistent results with the ground-truth abundance, while other methods generate false estimation on the middle part and lose the estimation on the right part of the ground-truth abundance map.

Compared with other NMF methods, our proposed SpNMFB and SpNMFP methods achieve much robust unmixing performance by adaptively assigning small weights to noisy bands and pixels according to SPL. Fig. 14 shows two noisy bands (i.e., bands 2 and 224), and Fig. 15 shows the the weights of different bands estimated by SpNMFB. It can be seen that the noisy bands, such as bands 1–3, 108–112, 154–166, and 220–224, are assigned small weights. Therefore, SpNMFB can effectively suppress the negative effects of noisy bands and is much more robust to band noise.

Fig. 16 shows the weights for different pixels estimated by SpNMFP. It can be seen that the pixels on the edge of objects or on the regions with much detailed information have large

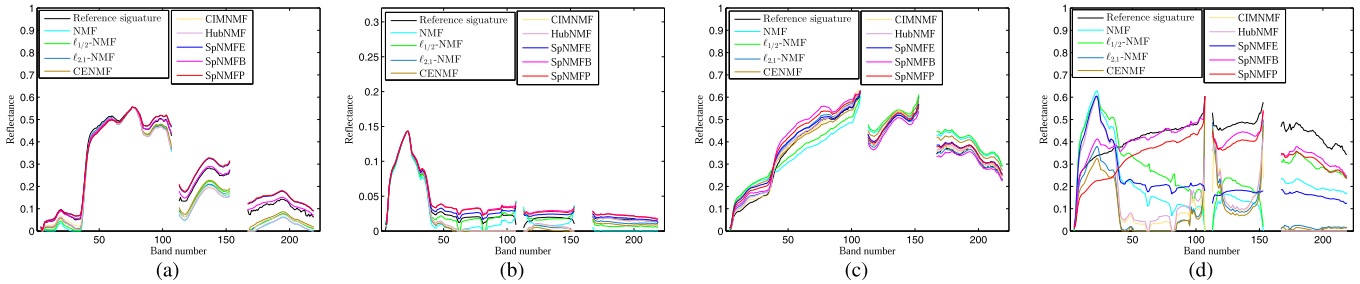


Fig. 18. Endmember signatures estimated by different methods over the Jasper data without the noisy bands. (a) Tree. (b) Soil. (c) Water. (d) Road.

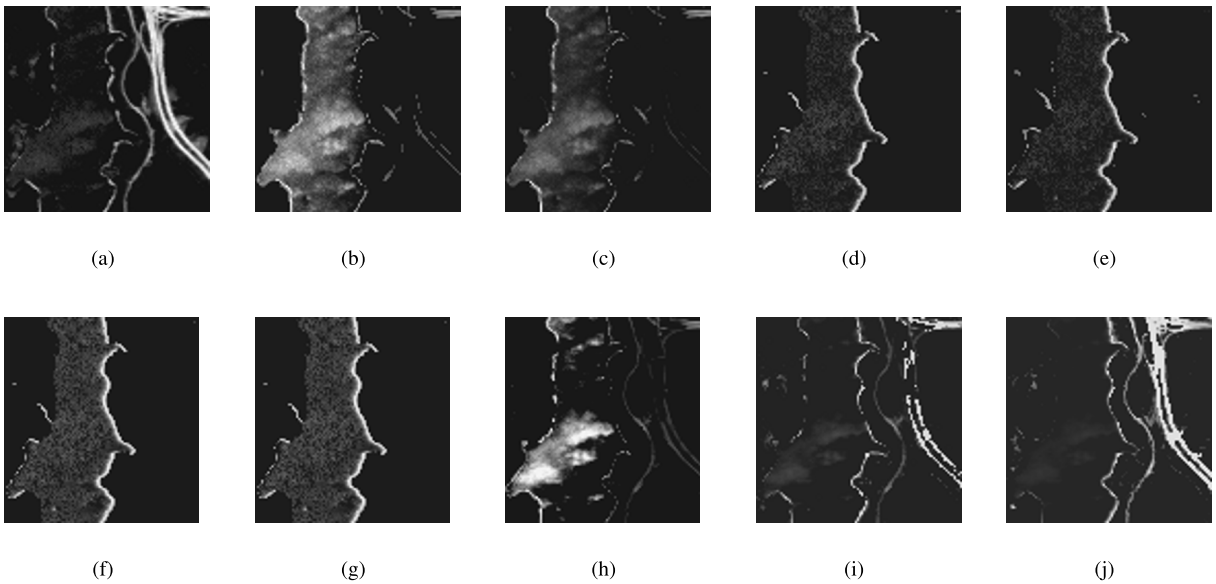


Fig. 19. Abundance maps for endmember 4 ("Road") of Jasper data without the noisy bands. (a) Ground truth. (b) NMF. (c) $\ell_{1/2}$-NMF. (d) $\ell_{2,1}$-NMF. (e) CENMF. (f) CIMNMF. (g) HubNMF. (h) SpNMFE. (i) SpNMFB. (j) SpNMFP.

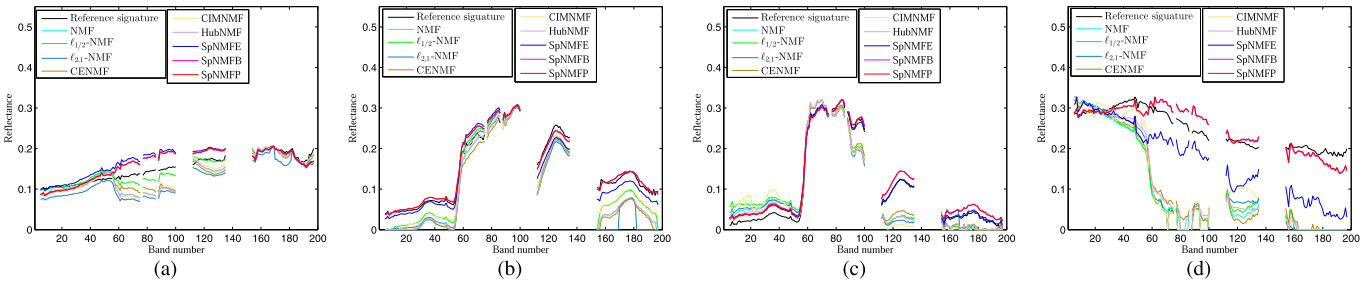


Fig. 20. Endmember signatures estimated by different methods over the Urban data. (a) Asphalt. (b) Grass. (c) Tree. (d) Roof.

TABLE IX
SAD OF DIFFERENT METHODS ON THE JASPER DATA SET WITHOUT THE NOISY BANDS

	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
Tree	0.2388	0.2123	0.2628	0.1522	0.2458	0.2145	0.1666	**0.0862**	0.0933
Water	0.3022	0.1119	0.2850	0.2882	0.2828	0.2990	**0.1081**	0.2050	0.1606
Soil	0.3739	0.1173	0.0909	0.1766	0.0831	0.2971	0.0960	0.0759	**0.0685**
Road	0.6143	0.5375	0.8346	0.8612	0.8430	0.7790	0.4731	0.2132	**0.1918**
Mean	0.3823	0.2447	0.3683	0.3695	0.3637	0.3974	0.2110	0.1451	**0.1285**

weights, while the pixels in the inner of homogenous regions have small weights. This means that not all pixels are useful for the unmixing, and selecting some representative "good" pixels can benefit the unmixing.

Fig. 17 shows the SAD results of SpNMFB and SpNMFP in the case of different numbers of repetitions. It can be seen that the performance of SpNMFB and SpNMFP can be improved by repeatedly executing the algorithms several times,

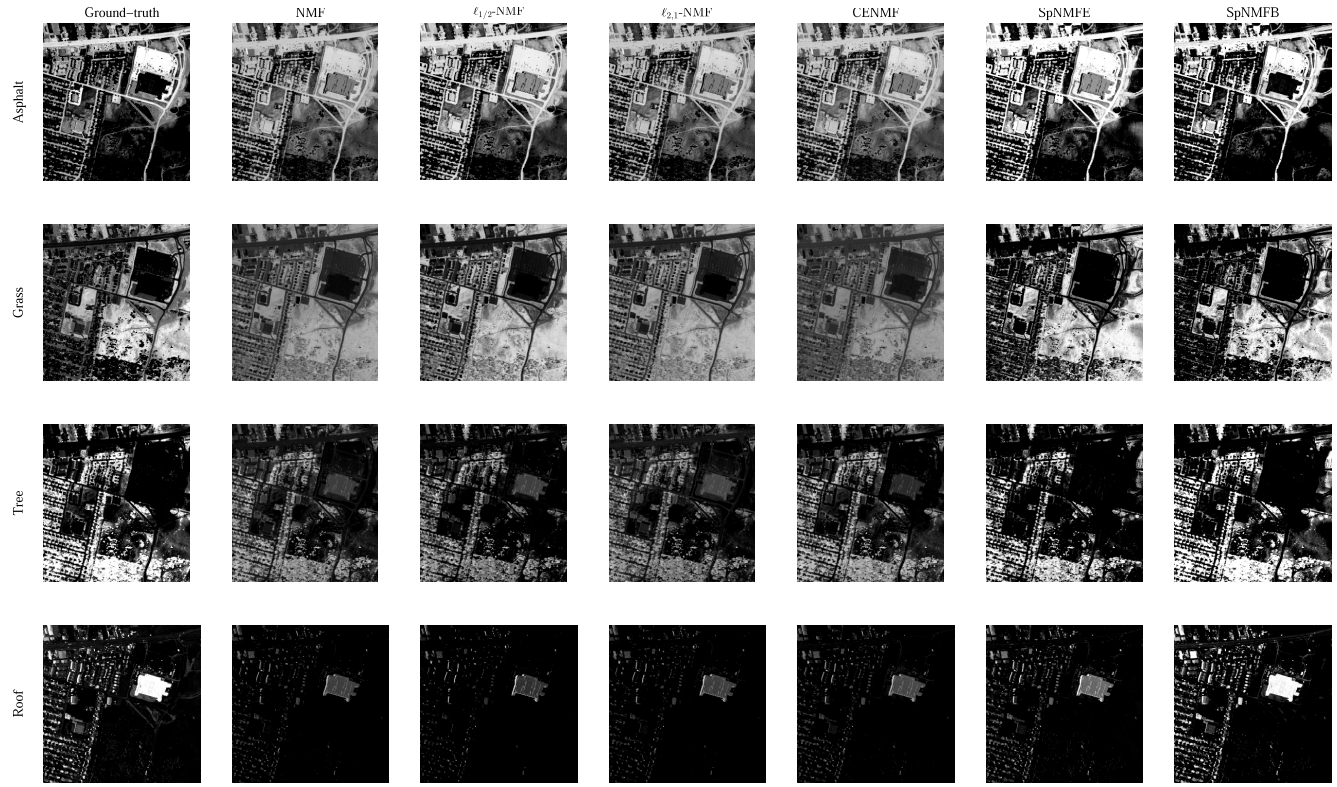


Fig. 21. Abundance maps on the Urban data. Each row denotes the maps of ground truth, NMF, $\ell_{1/2}$-NMF, $\ell_{2,1}$-NMF, CENMF, SpNMFE, and SpNMFB, respectively.

TABLE X
SAD OF DIFFERENT METHODS ON THE URBAN DATA SET

	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CIM-NMF	Huber-NMF	SpNMFE	SpNMFB	SpNMFP
Asphalt	0.1643	**0.1050**	0.1728	0.1474	0.1763	0.1762	0.1458	0.1109	0.1124
Grass	0.3867	0.3044	0.4122	0.2952	0.2986	0.3027	0.0992	**0.0532**	0.0543
Tree	0.2318	0.1885	0.2459	0.2304	0.2597	0.2611	**0.0871**	0.0986	0.1001
Roof	0.6751	0.6457	0.6922	0.6538	0.6702	0.6561	0.3085	0.0916	**0.0907**
Mean	0.3645	0.3109	0.3808	0.3317	0.3512	0.3490	0.1602	**0.0886**	0.0894

and the number of repetitions $R = 10$ corresponds to good results.

Table IX shows the SAD results obtained by different methods on the Jasper data set without noisy bands (i.e., 198 bands). Fig. 18 shows the estimated endmember signatures, and Fig. 19 shows the estimated abundance maps corresponding to material "Road." The results on 198 bands are similar to the abovementioned results on the 224 bands. By deleting noisy bands in advance, our proposed SpNMF methods also show relatively better results than other methods. Compared with the results on the full bands, on endmember "Road," the endmember signature and abundance map of our proposed SpNMF methods (i.e., SpNMFE, SpNMFB, and SpNMFP) are slightly improved, while those of other comparison methods even become worse by deleting noisy bands.

The second data set used in the experiment is the Urban data set. The spectral and spatial resolutions are 10 nm and 2 m, respectively. The scene has 307 $\times$ 307 pixels and 210 wavelengths ranging from 400 to 2500 nm. Due to the dense water vapor and atmospheric effects, the channels 1–4, 76, 87, 101–111, 136–153, and 198–210 are removed, and the remaining 162 bands are used for hyperspectral unmixing analyses [5], [8]. There are four endmembers: Asphalt, Grass, Tree, and Roof.

Table X shows the SAD results obtained by different methods on the urban data set. Fig. 20 shows the estimated endmember signatures by different methods. It can be seen that our proposed SpNMF methods are much more accurate than other methods especially on the materials: Grass, Tree, and Roof. Taking "Roof" as an example, the estimated endmember signatures of other methods are distorted, while our methods still can approximate the true endmember curve. Fig. 21 shows the estimated abundance maps by different methods. It is clear that our SpNMFB generates abundance maps that are consistent with the ground-truth maps.

Fig. 22 shows the SAD results of SpNMFB and SpNMFP in the case of different numbers of repetitions. It can be seen

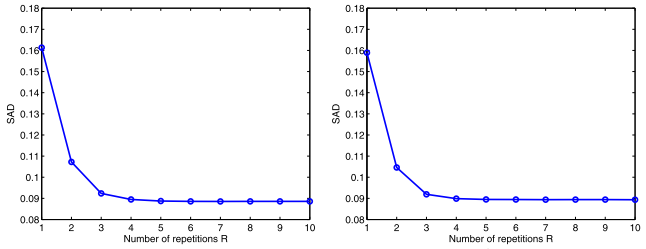


Fig. 22. SAD versus number of repetitions for (a) SpNMFB and (b) SpNMFP on the Urban data.

that SpNMFB and SpNMFP achieve good performance when the number of repetitions R is larger than 4.

V. Conclusion

In this article, we proposed three robust self-paced non-NMF (SpNMF) unmixing models to cope with the rowwise band noise, the columnwise pixel noise, and the elementwise image noise. The main characteristic of SpNMF is that an SPL strategy is used to automatically select appropriate atoms (bands or pixels or elements) for unmixing. By excluding bad atoms, the SpNMF models are more robust to noise and outliers. Experimental results on simulated and two real hyperspectral data sets have demonstrated that our proposed SpNMF methods are more accurate and robust than existing NMF methods. In particular, SpNMF methods can effectively discriminate noisy bands or pixels by assigning small weights for them.

References

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

[2] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, Oct. 1999.

[3] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[4] R. Huang, X. Li, and L. Zhao, "Spectral–spatial robust nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8235–8254, Oct. 2019.

[5] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[6] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[7] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 554–569, Apr. 2013.

[8] X.-R. Feng, H.-C. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6245–6257, Oct. 2018.

[9] B. Du, S. Wang, C. Xu, N. Wang, L. Zhang, and D. Tao, "Multi-task learning for blind source separation," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4219–4231, Sep. 2018.

[10] J. Peng and Q. Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.

[11] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.

[12] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy-based metric," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4027–4040, Nov. 2015.

[13] F. Zhu, A. Halimi, P. Honeine, B. Chen, and N. Zheng, "Correntropy maximization via ADMM: Application to robust hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4944–4955, Sep. 2017.

[14] Y. Ma, C. Li, X. Mei, C. Liu, and J. Ma, "Robust sparse hyperspectral unmixing with $\ell_{2,1}$ norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1227–1239, Mar. 2017.

[15] F. Zhu, Y. Wang, B. Fan, G. Meng, and C. Pan, "Effective spectral unmixing via robust representation and learning-based sparsity," 2014, *arXiv:1409.0685*. [Online]. Available: http://arxiv.org/abs/1409.0685

[16] D. Meng, Q. Zhao, and L. Jiang, "A theoretical understanding of self-paced learning," *Inf. Sci.*, vol. 414, pp. 319–328, Nov. 2017.

[17] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.

[18] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[19] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in *Proc. NIPS*, 2000, pp. 556–562.

[20] P. O. Hoyer, "Non-negative sparse coding," in *Proc. 12th IEEE Workshop Neural Netw. Signal Process.*, Sep. 2002, pp. 557–565.

[21] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. ICML*, 2009, pp. 41–48.

[22] Y. Jiang, D. Meng, Q. Zhao, S. Shan, and A G. Hauptmann, "Self-paced curriculum learning," in *Proc. AAAI*, 2015, pp. 2694–2700.

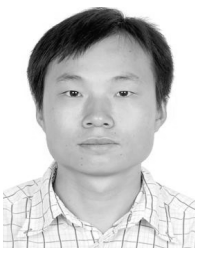
[23] Y.-D. Kim and S. Choi, "Weighted nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1541–1544.

[24] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using L21-norm," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2011, pp. 673–682.

[25] L. Du, X. Li, and Y.-D. Shen, "Robust nonnegative matrix factorization via half-quadratic minimization," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 201–210.

[26] X. Zhu and Z. Zhang, "Improved self-paced learning framework for nonnegative matrix factorization," *Pattern Recognit. Lett.*, vol. 97, pp. 1–7, Oct. 2017.

[27] Q. Zhao, D. Meng, L. Jiang, Q. Xie, Z. Xu, and A. G. Hauptmann, "Self-paced learning for matrix factorization," in *Proc. AAAI*, 2015, pp. 3196–3202.



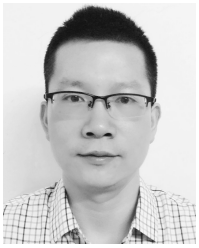
Jiangtao Peng received the B.S. and M.S. degrees from Hubei University, Wuhan, China, in 2005 and 2008, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2011.

He is a Professor with the Faculty of Mathematics and Statistics, Hubei University. His research interests include machine learning and hyperspectral image processing.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, all in electrical engineering.

He is an Associate Professor and the Director of the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou is a Senior Member of the International Society for Optical Engineering (SPIE). He was a recipient of the Third Prize of Macau Natural Science Award in 2014. He is a Co-Chair of the Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society. He serves as an Associate Editor for the IEEE TRANSACTIONS ON NEUTRAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and four other journals.



Weiwei Sun (Member, IEEE) received the B.S. degree in surveying and mapping and the Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2007, and 2013, respectively.

From 2011 to 2012, he studied with the Department of Applied Mathematics, University of Maryland College Park, College Park, MD, USA, working as a Visiting Scholar with Prof. J. Benedetto studying the dimensionality reduction of hyperspectral image. From 2014 to 2016, he studied with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, working as a Post-Doctoral Researcher to study intelligent processing in Hyperspectral imagery. From 2017 to 2018, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. He is a Full Professor with Ningbo University, Ningbo, China. He has authored or coauthored more than 70 journal articles. His research interests include hyperspectral image processing with manifold learning, anomaly detection and target recognition of remote sensing imagery using compressive sensing.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2000.

She is a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a fellow of the SPIE-International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She served as a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair with the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition from 2010 to 2014. She was a General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), the *Journal of Applied Remote Sensing*, and IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.

Lekang Xia is an undergraduate student with the Faculty of Mathematics and Statistics, Hubei University, Wuhan, China.