Self-Supervised Robust Deep Matrix Factorization for Hyperspectral Unmixing

Heng-Chao Li, Senior Member, IEEE, Xin-Ru Feng, Fellow, IEEE, Dong-Hai Zhai, Member, IEEE, Qian Du, Fellow, IEEE, and Antonio Plaza, Fellow, IEEE

Abstract—Hyperspectral unmixing is a critical step to process hyperspectral images (HSIs). Nonnegative matrix factorization (NMF) has drawn extensive attention in remotely sensed hyperspectral unmixing since it does not require prior knowledge about the pure spectral constituents (endmembers) in the scene. However, this approach is normally implemented as a single-layer procedure, which does not allow for a refinement of the obtained endmember abundances. In addition, HSIs suffer from the interference of sparse noise (besides Gaussian noise), which brings challenges when pursuing efficient hyperspectral unmixing. To address these issues, we propose a new self-supervised robust deep matrix factorization (SSRDMF) model for hyperspectral unmixing, which consists of two parts: encoder and decoder. In the encoder, a multilayer nonlinear structure is designed to directly map the observed HSI data to the corresponding abundances. The abundances are then decoded by the decoder, in which the connected weights are treated as the extracted endmembers. By modeling the sparse noise explicitly, the proposed method can reduce the effect caused by both Gaussian and sparse noise. Furthermore, a self-supervised constraint is included for exploring the spectral information, which is beneficial to further improve unmixing performance. To validate our method, we have conducted extensive experiments on both synthetic and real datasets. Our experiments reveal that our newly developed SSRDMF achieves superior unmixing performance compared to other state-of-the-art methods.

Index Terms—Deep matrix factorization, hyperspectral unmixing, self-supervised constraint, sparse noise.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) record the spectra of ground materials in hundreds of contiguous narrow bands [1]. Due to their wealth in terms of spectral information, HSIs have been applied to a variety of fields, such as food quality and safety assessment [2], precision agriculture [3], and road segmentation [4], among others. Due to the relatively low spatial resolution and complicated ground covers, the spectrum of a pixel in an HSI is inevitably composed of several materials, thus generating “mixed pixels.” To decrease the negative impact of such pixels on subsequent applications, hyperspectral unmixing has become an extensively used technique to decompose a mixed pixel into a collection of constituent materials (called endmembers) and their corresponding proportions (called abundances) [5].

Mixing models can be generally classified into linear mixture model (LMM) and nonlinear mixture model (NLMM). The former assumes a checker-board type macroscopic mixture, and an observed pixel is obtained by linearly mixing the endmembers with their associated abundances [6]. The latter holds when the mixing scale is microscopic, and multiple reflections and transmissions happen at the subpixel level [7], [8]. In view of its clear physical significance, simplicity, and effectiveness, the LMM has drawn considerable attention to address the hyperspectral unmixing problems [9], [10]. Under the LMM, three main classes of techniques are geometrical, sparse regression, and statistical-based ones [5]. Geometrical techniques exploit the fact that the mixed pixels are confined within a simplex under the linear mixture assumption, and the vertices of this simplex correspond to endmembers. Several endmember extraction algorithms have been developed based on the pure pixel assumption, such as N-FINDR [11] or vertex component analysis (VCA) [12]. To eliminate the need to have at least one pure pixel per endmember, minimum volume-based algorithms have also been developed [13], [14]. However, these methods are incapable of providing satisfactory results when dealing with highly mixed data. Techniques based on sparse regression are developed to seek for an optimal subset as endmembers from a spectral library [15], [16]. As a result, the accuracy of unmixing is closely related to the constructed library, which often cannot be adapted to images with different imaging conditions. In contrast, statistical techniques explore the statistical properties of HSIs and treat the unmixing problem as a blind source separation (BSS) one. A representative technique in this category is nonnegative matrix factorization (NMF) [17].

NMF can simultaneously extract endmembers and their associated abundance fractions by decomposing the data matrix under nonnegativity constraint. Nevertheless, nonconvexity of the cost function may result in local optimal solutions. To improve the unmixing performance, additional constraints exploiting the characteristics of HSIs have been
imposed on the endmembers [18]–[20], abundances [21]–[27], [29]–[32], and both [33]–[35]. For instance, the minimum volume constrained NMF (MVC-NMF) [18], minimum dispersion constrained NMF (MiniDisCo) [19], and endmember dissimilarity constrained NMF (EDCNMF) [20] regularize the endmember matrix. Liu et al. [21] proposed an abundance separation and smoothness constrained NMF (ASSNMF). Recently, sparsity-constrained unmixing methods have been used to minimize the number of endmembers contained in each mixed pixel. For instance, Yang et al. [22] proposed a gradient-based sparse NMF algorithm (NMF-SMC), where a nonlinear activation function between the layers to describe the nonlinearities. However, it is inconvenient to determine the final endmember and abundance matrices when the nonlinearity activation function is directly added to the factorized two matrices of each layer. To this end, inspired by the success of DL and autoencoder [44]–[47], we design a deep matrix factorization model with a multilayer nonlinear strategy to unmix hyperspectral data. In real applications, HSIs are inevitably contaminated by a mixture of the Gaussian noise and the sparse noise, where sparse noise refers to the noise with an arbitrary magnitude that affects several bands or pixels, caused by limited energy and poor imaging conditions [48], [49]. To relieve the impact of this kind of noise, an additional noise matrix is introduced to further construct a new robust deep matrix factorization (RDMF) model. Besides, in order to mine semantic information, self-supervised learning has been applied in [50]–[53]. In this article, we propose a new self-supervised RDMF (SSRDMF) method that reduces the negative influence caused by the Gaussian noise and the sparse noise, and better identifies endmember signatures and abundances from HSIs. The self-supervised constraint is imposed on the endmember matrix during our unmixing process for adequately exploiting the spectral information contained in HSIs. Specifically, candidate endmembers, i.e., multiple signatures within each class [54]–[56], are extracted automatically, and then, they are grouped by a fuzzy clustering to guide the learning process involving the endmembers. Our main contributions can be summarized as follows.

1) We introduce a new RDMF model with an asymmetric encoder–decoder framework for hyperspectral unmixing, which effectively addresses the complexity of HSIs and reduces the impact of Gaussian and sparse noise.

2) In the encoder part, a multilayer nonlinear network is designed to powerfully encode the original data, thus learning better abundances. The resulting abundances are then decoded by the decoder part with one layer, which is physically consistent with an extended LMM. In detail, an additional matrix is introduced to model the sparse noise explicitly according to the data acquisition process.

3) We further apply a self-supervised constraint for efficiently exploiting the spectral information in HSIs. In particular, a number of endmember candidates are extracted automatically, and then, they are grouped by a fuzzy clustering to guide the learning process for the endmembers.

4) Our new unmixing model adopts an iterative optimization algorithm based on gradient descent. We demonstrate that this algorithm leads to high unmixing accuracy using both synthetic and real datasets.

The remainder of this article is organized as follows. In Section II, we briefly describe some related works. Section III presents the proposed SSRDMF algorithm in detail, along with its optimization and implementation. Section IV describes our experiments and discusses the obtained results. Section V concludes this article with some remarks.

II. RELATED WORK

A. Linear Mixture Model

Let an HSI with $B$ bands and $P$ pixels be denoted as $X \in \mathbb{R}^{B \times P}$. Let $A \in \mathbb{R}^{B \times M}$ and $S \in \mathbb{R}^{M \times P}$ refer to the
spectral signatures with $M$ endmembers and abundance maps, respectively. The mixed data can be expressed as

$$X = AS + N$$  \hspace{1cm} (1)$$
where $N \in \mathbb{R}^{B \times P}$ stands for the Gaussian noise matrix. The abundance matrix $S$ satisfies the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC), i.e., $S_{mp} \geq 0$ and $\sum_m S_{mp} = 1$, $m = 1, 2, \ldots, M$ and $p = 1, 2, \ldots, P$.

To reduce the adverse impact of sparse noise, a sparsity regularized robust NMF ($L_{1/2}$-RNMF) is proposed in [48], given as

$$X = AS + E + N$$  \hspace{1cm} (2)$$
where $E$ describes the sparse noise using a $B \times P$ matrix.

B. Nonnegative Matrix Factorization and Its Variants

NMF [17] aims at decomposing an original nonnegative matrix into the product of two nonnegative matrices, which can be represented as $X \approx AS$. However, many pairs $(A, S)$ satisfy this approximation, using $AS = (A\Phi)(\Phi^{-1}S)$ with any invertible matrix $\Phi$. To solve $A$ and $S$, a cost function based on the Frobenius norm is first given to measure the approximation between $X$ and $AS$, i.e.,

$$C = \frac{1}{2} \|X - AS\|_F^2$$  \hspace{1cm} (3)$$
where the operator $\| \cdot \|_F^2$ denotes the Frobenius norm. Although model (3) is convex in $A$ and $S$ independently, it is nonconvex in $A$ and $S$ together. Therefore, a local optimal solution to model (3) can only be obtained by using optimization methods, such as gradient descent (GD) and half-quadratic programming. In general, according to Lee and Seung [17], multiplicative update rules are deduced as

$$A \leftarrow A \odot (XS^T) \odot (ASS^T)$$  \hspace{1cm} (4a)$$
$$S \leftarrow S \odot (A^TX) \odot (A^TAS)$$  \hspace{1cm} (4b)$$

where $\odot$ and $\odot$ stand for the elementwise multiplication and division, respectively, and $(\cdot)^T$ denotes the transpose operation to a matrix.

In order to improve the uniqueness of solution pairs $(A, S)$ by incorporating additional constraints on $A$ and $S$, the NMF-based hyperspectral unmixing problem can be written as

$$C = \frac{1}{2} \|X - AS\|_F^2 + \alpha \phi(A) + \beta R(S)$$  \hspace{1cm} (5)$$
where $\phi(A)$ and $R(S)$ are regularization terms for endmembers and abundances, respectively. Nonnegative parameters $\alpha$ and $\beta$ balance the effect of the corresponding constraints, respectively, resulting in various unmixing algorithms.

1) By setting $\alpha \neq 0$ and $\beta = 0$, resulting algorithms incorporate the constraints on endmembers, such as MVC-NMF [18], MiniDisCo [19], and EDCNMF [20].

2) There are numerous reports about imposing different constraints on abundances, i.e., $\alpha = 0$ and $\beta \neq 0$. In particular, both the abundance separation constraint and the abundance smoothness constraint were adopted in [21].

The sparsity constraint has been adopted by SMC [22] and $L_{1/2}$-norm regularizer [23]. For the latter, $R(S) = \|S\|_{1/2}$. On this basis, other constraints have been further investigated, mainly including graph regularization [24], structure constraint [25], substance dependence constraint [26], hypergraph regularizer [27], and subspace structure [28]. Moreover, there are joint sparsity and low-rank constraints [29], reweighted sparsity and total variation [30], manifold regularization [31], and region-based structure constraint [32].

3) The problem (5) with $\alpha \neq 0$ and $\beta \neq 0$ is considered simultaneously on endmembers and abundances [33]–[35].

C. Multilayer/Deep Autoencoder-Like NMF

In [40], the basic NMF is extended into an MLNMF structure for hyperspectral unmixing. Concretely, in the first layer, $X$ is decomposed into $A_1$ and $S_1$. Then, the matrix $S_1$ acquired over the first layer is decomposed into $A_2$ and $S_2$. The same process is continued (i.e., $S_{l-1} \approx A_lS_l$) until the maximum number of layers $L$ is reached. The endmember and abundance matrices for the MLNMF can be formulated as

$$A = A_1A_2\cdots A_L, \quad S = S_L$$  \hspace{1cm} (6)$$
where $A_l \in \mathbb{R}^{M_l \times M_l}, l = 1, 2, \ldots, L$, and $M_0$ equals to the number of bands $B$.

In [57], the unified cost function of the deep autoencoder-like NMF model is given as

$$C = \|X - A_1A_2\cdots A_LS_L\|_F^2 + \|S_L - A_L^T\cdots A_2^TA_1^TX\|_F^2 + \beta \text{Tr}(S_LGS_L^T)$$  \hspace{1cm} (7)$$
\text{s.t. } S_L \geq 0, \quad A_l \geq 0, \quad l = 1, 2, \ldots, L$$

where $A_l$ is the weight factor of the $l$th layer, $S_L$ is a so-called community membership matrix, $G$ represents the graph Laplacian matrix, and $\beta$ is a parameter for the graph regularizer.

III. PROPOSED METHOD

A. Proposed SSRDMF Model

In [57], a deep autoencoder-like NMF is provided to learn hierarchical features with hidden information, whose factor matrices are however decomposed linearly layer by layer. Moreover, it shares the information from complex HSI data contaminated by Gaussian and sparse noise, we design an RDMF model that consists of two asymmetric parts: encoder and decoder. That is, the image is encoded using a multilayer nonlinear strategy, whose output is decoded by a one-layer decoder. Due to the complexity of the observation data, a multilayer nonlinear structure for
Encoder is built to directly map $X$ into the abundance space, thus improving the ability to encode $X$. Specifically, $S_1$ is learned from $X$ in the first layer. Similarly, $S_2$ is learned from $S_1$ in the next layer. The same process is continued until the $S_L$ is learned. The formulation of each layer is written into

$$S_1 = \sigma(W_1 X)$$
$$S_2 = \sigma(W_2 S_1)$$
$$\vdots$$
$$S_L = \sigma(W_L S_{L-1})$$

(8)

where $\sigma(\cdot)$ is a nonlinear activation function, $S_l$ is the output matrix of the $l$th layer, and $W_l$ is a weight matrix connecting the $(l-1)$th layer with the $l$th layer. As such, by integrating all layers, (8) becomes

$$S_L = \sigma(W_L \cdots \sigma(W_2 \sigma(W_1 X)))$$

(9)

After obtaining the improved abundances, a decoder is incorporated into the unmixing process to reconstruct the image and obtain the endmembers. Furthermore, it is worth noting that $W_l$ is not associated with the weight of the decoder and does not have to satisfy nonnegativity. Consequently, it is reasonable that the number of layers is different. In particular, a one-layer decoder is applied for hyperspectral unmixing, being physically consistent with the extended LMM given in (2). Particularly, an additional matrix $E$ is introduced in the decoder to characterize sparse noise. By integrating the decoder and the encoder into a unified framework, the cost function of the proposed RDMF model is given as

$$C = \frac{1}{2P} \|X - E - A \sigma(W_L \cdots \sigma(W_2 \sigma(W_1 X)))\|^2_F$$
$$+ \frac{\gamma}{P} \|E\|_{1,2}, \text{ s.t. } A \geq 0, S_L \geq 0, \ 1^T_M S_L = 1^T_P$$

(10)

where $\gamma$ denotes a balance parameter to tune the row sparsity of $E$, $1_M$ is a column vector of all ones with $M$ elements, and $\|E\|_{1,2}$ is calculated by $\|E\|_{1,2} := \sum_{l=1}^L \|E_l\|_2$. By doing so, the proposed model (10) can alleviate the negative influence caused by Gaussian and sparse noise simultaneously. As a result, the output $S_L$ of the encoder acts as an estimation of the abundance matrix and the decoder weight $A$ as the extracted endmember matrix.

Taking the information of spectral signatures into account, the self-supervised constraint is imposed on the endmember matrix. First, candidate endmembers $\Xi \in \mathbb{R}^{B \times N}$ are extracted from the HSI by using existing endmember extraction methods, for which the $n$th column vector $\Xi_n$ denotes one candidate endmember. For example, by using the VCA algorithm $k_{VCA}$ times, there are $N = k_{VCA} \times M$ signatures in $\Xi$. Then, $\Xi$ is grouped by the FCM clustering, thus generating the cluster centroids $C$ to guide the learning process of endmember matrix $A$. The cost function of the proposed SSRDMF model is formed as

$$C = \frac{1}{2P} \|X - E - A \sigma(W_L \cdots \sigma(W_2 \sigma(W_1 X)))\|^2_F$$
$$+ \frac{\eta}{N} \sum_{n,m} u_{nm} \|\Xi_n - C_m\|^2_2 + \frac{\mu}{2P} \|A - C\|^2_F$$

s.t. $A \geq 0, S_L \geq 0, \ 1^T_M S_L = 1^T_P$

(11)

where $\alpha$ is a balance parameter that controls the strength of fuzzy clustering, $\eta \in [1, +\infty]$ represents the degree of fuzziness, $\mu$ is a penalty parameter, and $\|\cdot\|_2$ is $L_2$-norm. Each column $C_m$ in matrix $C$ denotes the vector of the centroid of a cluster. Meanwhile, $u_{nm}$ is the element in the membership matrix $U$, which represents the degree of membership of $\Xi_n$ in the cluster $m$, and it should satisfy the constraints: $0 \leq u_{nm} \leq 1$ and $\sum_{m=1}^M u_{nm} = 1$. A graphical illustration of SSRDMF is given in Fig. 1.

**B. Optimization**

It is obvious that (11) is nonconvex with respect to $A$, $E$, $C$, $U$, and $\{W_l\}_{l=1}^L$. As such, to successfully optimize the problem, the underlying variables require to be updated alternately. In order to facilitate the optimization, let $S_L = \sigma(W_L \cdots \sigma(W_2 \sigma(W_1 X)))$, and (11) can be further
transformed into the following cost function:
\[
C = \frac{1}{2P}\|X - E - \text{AS}_L\|_F^2 + \frac{\gamma}{P}\|E\|_{1,2}
+ \frac{\alpha}{N}\sum_{n,m} u_{nm}^\eta \|\Xi_n - C_m\|_2^2 + \frac{\mu}{2P}\|A - C\|_F^2
\]
\[\text{s.t. } A \geq 0, \ S_L \geq 0, \ 1^T S_L = 1^T. \quad (12)\]

As such, problem (12) is optimized by updating alternately \(A, E, C, U,\) and \(\{W_l\}_{l=1}^L\), whose update rules are developed in detail in the following.

1) Update Rule for \(A\): In order to ensure the nonnegativity of matrix \(A\), the GD method is utilized to obtain the update rule for \(A\). The derivative of (12) with respect to \(A\) is given as
\[
\frac{\partial C}{\partial A} = \frac{1}{P}\text{AS}_L S_L^T - \frac{1}{P}(X - E)S_L^T + \frac{\mu}{P} A - \frac{\mu}{P} C. \quad (13)
\]
By moving \(A\) in the opposite direction of the gradient with appropriate step size, \(A\) can be updated by
\[
A \leftarrow A \odot \left((X - E)S_L^T + \mu C\right) \odot (\text{AS}_L S_L^T + \mu A). \quad (14)
\]

2) Update Rule for \(E\): Recalling the cost function given in (12), \(E\) can be solved according to
\[
E \leftarrow \arg \min_{E} \frac{1}{2P}\|X - E - \text{AS}_L\|_F^2 + \frac{\gamma}{P}\|E\|_{1,2}. \quad (15)
\]

According to [48], the \(i\)th row of \(E\) can be formulated as
\[
E_{bi} = \begin{cases} \frac{\|Q_{bi} - \gamma Q_{bi}\|}{\|Q_{bi}\|} & \text{if } \|Q_{bi}\|_2 \geq \gamma \\ 0, & \text{otherwise} \end{cases} \quad (16)
\]
where \(Q = (X - \text{AS}_L)\), and 0 is an all-zero vector with size \(1 \times P\).

3) Update Rule for \(C\): Similarly, the derivative of (12) with respect to \(C_m\) is deduced as
\[
\frac{\partial C}{\partial C_m} = \frac{2\alpha}{N}\sum_n u_{nm}^\eta (C_m - \Xi_n) + \frac{\mu}{P}(C_m - A_m). \quad (17)
\]
Let \(\partial C/\partial C_m\) be set to 0. Then, \(C_m\) can be updated according to
\[
C_m \leftarrow \frac{2\alpha}{N}\sum_n u_{nm}^\eta \Xi_n + \frac{\mu}{P} A_m. \quad (18)
\]
It should be noted that the auxiliary variable \(C\) is updated in vector/row form.

4) Update Rule for \(U\): To relax the constraints of \(0 \leq u_{nm} \leq 1\) and \(\sum_{m=1}^M u_{nm} = 1\), the Lagrangian function of (12) with respect to \(U\) is represented as
\[
\mathcal{J} = \frac{\alpha}{N}\sum_{n,m} u_{nm}^\eta \|\Xi_n - C_m\|_2^2 + \frac{1}{N}\sum_n \lambda_n \left(\sum_m u_{nm} - 1\right)
\]
\[\text{where } \lambda_n \text{ is the Lagrangian multiplier. Let the derivation of} \quad (19)\text{with respect to } u_{nm} \text{be 0 and the update rule for } u_{nm} \text{be given by} \quad (58)\]
\[
u_{nm} = \frac{1}{\sum_{z} \left(\frac{\Xi_n - C_m}{\Xi_n - C_{\hat{m}}}\right)^2} \quad (20)
\]

### Algorithm 1 Proposed SSRDMF Model for Hyperspectral Unmixing

**Input:** Observation matrix \(X\);  
Number of endmembers \(M\);  
Tunable parameters \(\gamma, \alpha,\) and \(L\).

**Initialize:** \(A\) by VCA or random values from 0 to 1, \(E = 0, \ U\) by random initialization, \(\{W_l\}_{l=1}^L\) by the L-BFGS algorithm with (26b), \(\eta = 2, \delta = 15, T_{\text{max}} = 300,\) and \(\varepsilon = 1 \times 10^{-4}\).

**Repeat**

- Update \(A\) by (14),
- Update \(E\) by (16),
- Update \(C\) by (18),
- Update \(U\) by (20),
- Update \(\{W_l\}_{l=1}^L\) by L-BFGS with (22)-(24),

**until** stopping criterion is satisfied.

**Output:** Endmember signature matrix \(A\),  
Abundance matrix \(S_L\).

Specifically, if \(\|\Xi_n - C_m\|_2 \geq \eta\), then \(u_{nm} = 1\), and the membership degrees of the points in other classes will be set as 0.

5) Update Rules for \(\{W_l\}_{l=1}^L\): In order to calculate the derivative for the weight matrix \(W_l\) of the \(l\)th layer, we have the following chain rule:
\[
\frac{\partial C}{\partial (W_l)_{ki}} = \sum_j \frac{\partial C}{\partial (W_l S_{l-1})_{kj}} \frac{\partial (W_l S_{l-1})_{kj}}{\partial (W_l)_{ki}} = \left[\frac{\partial C}{\partial (W_l S_{l-1})} S_{l-1}^T\right]_{ki}. \quad (21)
\]
By combining \(S_l = \sigma(W_l S_{l-1})\) and using matrix notation of (21), the following update equation is obtained:
\[
\frac{\partial C}{\partial W_l} = \left[\frac{\partial C}{\partial S_l} \odot \nabla \sigma(W_l S_{l-1})\right] S_{l-1}^T \quad (22)
\]
for \(l = 1, 2, \ldots, L\), \(\nabla\) denotes the differential operator, and \(S_{l-1}\) is the output of the \((l-1)\)th layer for the forward propagation process, i.e., \(S_{l-1} = \sigma(W_{l-1} \cdots \sigma(W_1 X))\). It is clear that (22) includes the \(\partial C/\partial S_l\) term. Therefore, the derivative with respect to \(S_l\) is computed by
\[
\frac{\partial C}{\partial S_l} = W_{l+1}^T \left[\frac{\partial C}{\partial S_{l+1}} \odot \nabla \sigma(W_{l+1} S_l)\right] \quad (23)
\]
for \(l = 1, 2, \ldots, L - 1\). In particular, for the last layer (i.e., the \(L\)th layer), \(\partial C/\partial S_L\) can be computed as
\[
\frac{\partial C}{\partial S_L} = \frac{1}{P} A^T (\text{AS}_L - (X - E)). \quad (24)
\]
After computing the above derivatives with respect to \(\{W_l\}_{l=1}^L\), the optimization is implemented based on the limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method, which is a quasi-Newtonian method.

As a result, based upon the above update rules of all variables, the optimization of the proposed SSRDMF model is summarized in Algorithm 1.
To implement effectively the proposed unmixing method, several issues need to be considered.

1) The number of endmembers is a crucial factor for hyperspectral unmixing, whereas it remains an open topic. To this end, we set the number of endmembers manually or estimate it through an effective method, such as virtual dimensionality [59] and HySime [60].

2) The second issue is how to initialize the related variables, including \( \mathbf{A} \), \( \mathbf{E} \), \( \mathbf{U} \), and \( \{\mathbf{W}_l\}_{l=1}^L \). This process can be also regarded as pretraining stage from the perspective of DL. The endmember matrix \( \mathbf{A} \) is initialized by VCA [12] or random values from 0 to 1. Meanwhile, \( \mathbf{E} \) is initialized with a zero matrix. The membership matrix \( \mathbf{U} \) is randomly initialized. For the initialization of \( \{\mathbf{W}_l\}_{l=1}^L \), all factors require to be updated layer by layer. Specifically, for the first layer, the weight matrix \( \mathbf{W}_1 \) is initialized by adopting the basic autoencoder with one-layer structure from the given nonnegative matrix \( \mathbf{X} \). After obtaining \( \mathbf{W}_1 \), \( \mathbf{S}_1 \) can be calculated naturally by \( \mathbf{S}_1 = \sigma(\mathbf{W}_1 \mathbf{X}) \). Similarly, for the second layer, \( \mathbf{S}_1 \) is treated as the input of the autoencoder, and \( \mathbf{W}_2 \) can be learned. The successional initialization process is performed until \( \mathbf{W}_L \) is obtained. To sum up, from the first to the final layer, in order to learn the weight matrix of each layer \( \mathbf{W}_l \), the cost function based on Euclidean distance is formulated as

\[
C_{AE} = \frac{1}{2P} \| \mathbf{S}_l - \mathbf{H}_l \sigma(\mathbf{W}_l \mathbf{S}_{l-1}) \|^2_T, \text{ s.t. } \mathbf{H}_l \geq 0 \tag{25}
\]

where \( \mathbf{H}_l \) is the weight matrix connecting the hidden layer with the output layer and \( \mathbf{S}_l \) is the representation of the hidden layer in the autoencoder. Analogously, the derivatives of (25) with respect to \( \mathbf{H}_l \) and \( \mathbf{W}_l \) are calculated, respectively, as

\[
\frac{\partial C_{AE}}{\partial \mathbf{H}_l} = \frac{1}{P} \left[ \mathbf{H}_l \sigma(\mathbf{W}_l \mathbf{S}_{l-1}) - \mathbf{S}_{l-1} \right] [\sigma(\mathbf{W}_l \mathbf{S}_{l-1})]^T \tag{26a}
\]

\[
\frac{\partial C_{AE}}{\partial \mathbf{W}_l} = \frac{1}{P} \left[ \mathbf{H}_l^T (\mathbf{H}_l \sigma(\mathbf{W}_l \mathbf{S}_{l-1}) - \mathbf{S}_{l-1}) \right.
\]

\[
\left. \otimes \nabla \sigma(\mathbf{W}_l \mathbf{S}_{l-1}) \mathbf{S}^{T}_{l-1} \right] \tag{26b}
\]

Let \( \frac{\partial C_{AE}}{\partial \mathbf{H}_l} \) be set as 0. Then, the following update rule for \( \mathbf{H}_l \) can be obtained by:

\[
\mathbf{H}_l \leftarrow (\mathbf{S}_{l-1} [\sigma(\mathbf{W}_l \mathbf{S}_{l-1})]^T) [\sigma(\mathbf{W}_l \mathbf{S}_{l-1}) (\sigma(\mathbf{W}_l \mathbf{S}_{l-1}))^T]^{-1} \tag{27}
\]

where \( (\cdot)^{-1} \) denotes the inverse of a matrix. L-BFGS is also employed to update the variable \( \mathbf{W}_l \). The above variables \( \mathbf{H}_l \) and \( \mathbf{S}_l \) are initialized according to VCA and fully constrained least squares (FCLS), respectively.

3) The next issue involves the ANC and the ASC. The nonnegativity of the matrices \( \mathbf{A} \) and \( \{\mathbf{S}_l\}_{l=1}^L \) can be satisfied under the derivative update rules. Then, when \( \mathbf{S}_L \) is calculated, the ASC is enforced by redefining the augmented matrices as [48]

\[
\mathbf{X} - \mathbf{E} = \begin{bmatrix} \mathbf{X} - \mathbf{E} \\ \delta I_P \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A} \\ \delta I_M \end{bmatrix} \tag{28}
\]

where \( \delta \) controls the impact of the ASC.

4) As for the stopping criterion, our proposed method is terminated when the number of iterations reaches a preset value \( T_{\text{max}} \), or when the relative error \( C(t) - C(t+1) \leq \epsilon C(t) \) is satisfied, where \( t \) represents the \( t \)th iteration and \( \epsilon \) is the error threshold. \( T_{\text{max}} \) and \( \epsilon \) are set to be 300 and \( 1 \times 10^{-5} \), respectively. Furthermore, \( C \) is calculated according to (11) in Algorithm 1, while \( C \) equals \( C_{AE} \) given in (25) for initializing the matrices \( \{\mathbf{W}_l\}_{l=1}^L \).

IV. EXPERIMENTAL RESULTS

Experiments on synthetic and real hyperspectral datasets have been conducted to demonstrate the effectiveness of the proposed SSRDMF method. \( L_1 \)-DNMF [43], uDAS [45], MLNMF [40], \( L_{1/2} \)-RNMF [48], \( L_{1/2} \)-NMF [23], and VCA-FCLS [12], [61] are adopted for comparison. For the proposed method, the underlying parameters \( \delta \) and \( \eta \) are set to be 15 and 2, respectively. For the sake of a reliable comparison, all compared methods are initialized by using the VCA and FCLS algorithm, all experiments run ten times, and the average values along with their standard deviations are reported. Moreover, all experiments are conducted under the environment of MATLAB R2015b software and computer configuration Intel Core i5 CPU @ 2.80 GHz and 8.00-GB RAM.

Two metrics are utilized to assess unmixing performance quantitatively: the spectral angle distance (SAD) and the root mean square error (RMSE), given as

\[
\text{SAD}_m = \arccos \left( \frac{\mathbf{A}_m^T \hat{\mathbf{A}}_m}{\|\mathbf{A}_m\| \|\hat{\mathbf{A}}_m\|} \right) \tag{29a}
\]

\[
\text{RMSE}_m = \sqrt{\frac{1}{P} \| \mathbf{S}_m - \hat{\mathbf{S}}_m \|^2} \tag{29b}
\]

where \( \mathbf{A}_m \) and \( \hat{\mathbf{A}}_m \) are the \( m \)th original and estimated endmember spectral signatures, and \( \mathbf{S}_m \) and \( \hat{\mathbf{S}}_m \) are the \( m \)th original and estimated abundances, respectively.

A. Synthetic Data Experiments

Synthetic data are generated by using the United States Geological Survey (USGS) spectral library, which is available online at http://speclab.cr.usgs.gov/spectral.lib06. It comprises nearly 500 typical minerals, and each spectral signature is given in 224 bands ranging from 0.4 to 2.5 \( \mu \)m. In this article,
eight spectral signatures are selected as the endmembers (shown in Fig. 2), and their abundance values are generated based on LMM following the method in [9]. The synthetic image is constructed with $64 \times 64$ pixels and 224 bands per pixel. We then add the Gaussian noise and the sparse noise. For the former, we use different noise levels in terms of the signal-to-noise ratio (SNR). For the latter, impulse noise is added into the synthetic dataset, where ratio and sp are used to denote the percentage of the synthetic image bands corrupted by impulse noise and that of pixels for each corrupted band, respectively.

1) Experiment 1 (Analysis of the Number of Layers): For our proposed model, the number of layers is a key parameter. Thus, the number of layers $L$ is assigned from 1 to 6 with $M = 6$ when $\gamma = 5.0$ and $\alpha = 0.2$. A comparison of the obtained results in terms of SAD and RMSE is shown in Fig. 3. It is obvious that the curves plotted by these two metrics (for different noise levels) exhibit a similar trend. In detail, when $L$ equals 1, the proposed model degrades into the one-layer autoencoder, which achieves poor unmixing performance with respect to SAD and RMSE. The curves descend as $L$ grows in the beginning and then slightly rise with the further increase of $L$, implying that the deep architecture offers better performance gains over the one-layered structure. Overall, the proposed SSRDMF produces desirable results when $L$ is set as 3.

2) Experiment 2 (Parameter Analysis): In order to demonstrate the influence of parameters $\gamma$ and $\alpha$, experiments are conducted on the synthetic dataset by fixing SNR = 20 dB, ratio = sp = 0.05, and $M = 6$. $\gamma$ is chosen from the set $\{0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0, 15.0\}$, and $\alpha$ is tuned in finite set $\{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0\}$. Fig. 4(a) and (b), respectively, shows the SAD and RMSE results achieved by SSRDMF with respect to parameters $\gamma$ and $\alpha$. It can be observed that both SAD and RMSE change slightly when $\gamma$ is larger than 2.0. In short, the optimal result can be acquired when $\gamma = 4.0$ and $\alpha = 0.05$. Besides, a lower level of sparse noise is also associated with a larger value of $\gamma$. In fact, the optimal values of these parameters are relevant to other factors including (but not limited to) the number of endmembers.

3) Experiment 3 (Robustness to Gaussian Noise): In this experiment, the HSI data are contaminated by four Gaussian noise levels, i.e., 15, 20, 25, and 30 dB under ratio = sp = 0.10 and $M = 6$. The obtained results in terms of SAD and RMSE are shown in Table I. It is evident that all methods improve compared with VCA-FCLS. Overall, the performance of each method improves with the decrease in the Gaussian noise. Most importantly, our proposed method achieves better unmixing performance than the others. This indicates that SSRDMF is robust to the Gaussian noise.

4) Experiment 4 (Robustness to Sparse Noise): We also test the robustness of the proposed algorithm to sparse noise under $M = 6$. The Gaussian noise level is fixed to 25 dB. The sparse noise intensities vary from 0.05 to 0.20, with a stepsize increment of 0.05. Table II presents a comparison of the obtained results in terms of SAD and RMSE. It is obvious that the optimal results are generally acquired for lower degrees of sparse noise contamination. Particularly, SSRDMF obviously
outperforms other methods under high levels of sparse noise, revealing the importance of modeling sparse noise explicitly in our model. Analogously, SSRDMF yields the smaller average SAD and its standard deviation under each condition, which demonstrates its superiority and robustness.

5) Experiment 5 (Generalization to Different Numbers of Endmembers): In this experiment, different scenarios with various numbers of endmembers have been considered. Specifically, $M$ is set to 4, 6, 7, and 8. These images are degraded by the Gaussian noise with $\text{SNR} = 25 \text{ dB}$ and the sparse noise with ratio $= sp = 0.05$. As reported in Table III, better performance can be achieved with a lower number of endmembers for all methods. This may be due to the fact that the data simulated with fewer endmembers convey a more intrinsically simple structure. Besides, our proposed method provides smaller values of SAD and RMSE (especially for more endmembers), suggesting the effectiveness of our proposed model. Notably, obvious advantages can be found under high levels of the Gaussian noise and the sparse noise and more endmembers. This implies that our proposed SSRDMF can handle more complex HSIs.

6) Experiment 6 (Analysis of the Clustering Effect): To analyze the clustering effect, experiments are conducted on the synthetic dataset by fixing $\text{SNR} = 20 \text{ dB}$, ratio $= sp = 0.05$, and $M = 6$. We set $k_{\text{VCA}}$ as $\{1, 5, 10, 20, 30, 40, 50\}$ under $\gamma = 4.0$ and $\alpha = 0.05$. Here, for $k_{\text{VCA}} = 1$, $\Xi$ is equal to $C$, and the elements of the membership matrix are all one, i.e., without clustering. As far as the clustering is concerned, it should be noted that the cluster number is always equal to $M$ no matter how $k_{\text{VCA}}$ changes. The unmixing performance
TABLE IV

<table>
<thead>
<tr>
<th>Method</th>
<th>SSRDMF</th>
<th>$L_1$-DNMF</th>
<th>uDAS</th>
<th>MLNMF</th>
<th>$L_{1/2}$-RNMF</th>
<th>$L_{1/2}$-NMF</th>
<th>VCA-FCLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M=4$, SNR=25, ratio &amp; sp=0.05</td>
<td>78.1</td>
<td>131.1</td>
<td>4.5</td>
<td>7.9</td>
<td>18.9</td>
<td>2.8</td>
<td>0.7</td>
</tr>
<tr>
<td>$M=6$, SNR=20, ratio &amp; sp=0.05</td>
<td>127.9</td>
<td>145.7</td>
<td>5.7</td>
<td>9.4</td>
<td>19.8</td>
<td>1.6</td>
<td>1.3</td>
</tr>
<tr>
<td>$M=6$, SNR=25, ratio &amp; sp=0.05</td>
<td>132.7</td>
<td>191.2</td>
<td>5.6</td>
<td>9.2</td>
<td>19.9</td>
<td>2.3</td>
<td>1.3</td>
</tr>
<tr>
<td>$M=6$, SNR=25, ratio &amp; sp=0.10</td>
<td>126.4</td>
<td>200.1</td>
<td>34.4</td>
<td>9.7</td>
<td>19.8</td>
<td>1.8</td>
<td>1.3</td>
</tr>
<tr>
<td>$M=8$, SNR=25, ratio &amp; sp=0.05</td>
<td>268.0</td>
<td>362.8</td>
<td>91.2</td>
<td>11.9</td>
<td>20.7</td>
<td>3.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

TABLE V

<table>
<thead>
<tr>
<th>Methods</th>
<th>SAD (Average of Ten Runs)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boldfaced Number Denotes the Best Result Under Each Condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alumite GDS82 Na82</td>
<td>0.1165±0.272%</td>
<td>0.1682±0.217%</td>
</tr>
<tr>
<td>Andradite WS487</td>
<td>0.1289±1.884%</td>
<td>0.0932±0.293%</td>
</tr>
<tr>
<td>Buddingtonite GDS85 D-206</td>
<td>0.1219±2.195%</td>
<td>0.1133±0.985%</td>
</tr>
<tr>
<td>Chaledony CU91-6A</td>
<td>0.1269±1.866%</td>
<td>0.1392±2.400%</td>
</tr>
<tr>
<td>Kaolin/Snecet H89-FR-5 30K</td>
<td>0.1128±1.966%</td>
<td>0.1349±2.400%</td>
</tr>
<tr>
<td>Kaolinite Kga-2</td>
<td>0.1248±1.996%</td>
<td>0.1412±1.966%</td>
</tr>
<tr>
<td>Montmorillonite+Illi CM37</td>
<td>0.1252±1.106%</td>
<td>0.1374±2.180%</td>
</tr>
<tr>
<td>Muscovite IIL07</td>
<td>0.1252±1.106%</td>
<td>0.1374±2.180%</td>
</tr>
<tr>
<td>Nontronite NG-1.a</td>
<td>0.1033±1.478%</td>
<td>0.1300±1.136%</td>
</tr>
<tr>
<td>Pyrope WS474</td>
<td>0.0721±0.678%</td>
<td>0.0080±0.345%</td>
</tr>
<tr>
<td>Spheine HS189 3B</td>
<td>0.0668±0.848%</td>
<td>0.0216±0.444%</td>
</tr>
</tbody>
</table>

**Mean** | 0.0991±0.533% | 0.1071±0.545% | 0.1033±0.511% | 0.1024±0.579% | 0.1015±0.611% | 0.1083±0.811% | 0.1108±0.487% | 0.1165±0.664% | 0.1096±0.929% |

7) Experiment 7 (Convergence Observation): In order to illustrate the convergence of SSRDMF, we draw convergence curves of the value of $\mathcal{C}$ given in (11) on synthetic dataset contaminated by four Gaussian noise levels (i.e., 15, 20, 25, and 30 dB) under ratio = sp = 0.10 and $M = 6$, as shown in Fig. 6. From Fig. 6, as the number of iterations increases, all curves are monotonically decreasing and approach convergence although the number of stopping iterations is slightly different.
TABLE VI
SAD and RMSE scores (average of ten runs) along with their standard deviation on the Samson dataset for different methods. Boldfaced number denotes the best result under each condition.

<table>
<thead>
<tr>
<th>SAD</th>
<th>SSRDMF</th>
<th>L₁-DNMF</th>
<th>uDAS</th>
<th>MLNMF</th>
<th>L₁/²-RNMF</th>
<th>L₁/²-NMF</th>
<th>VCA-FCLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil</td>
<td>0.0520±0.94 %</td>
<td>0.0200±0.108 %</td>
<td>0.0259±0.73 %</td>
<td>0.0754±16.39 %</td>
<td>0.0281±0.39 %</td>
<td>0.0620±11.28 %</td>
<td>0.1546±24.32 %</td>
</tr>
<tr>
<td>Tree</td>
<td>0.0376±0.32 %</td>
<td>0.0328±0.17 %</td>
<td>0.0428±0.61 %</td>
<td>0.0590±3.87 %</td>
<td>0.0514±0.30 %</td>
<td>0.0647±5.03 %</td>
<td>0.0486±29.2 %</td>
</tr>
<tr>
<td>Water</td>
<td>0.0903±1.52 %</td>
<td>0.1314±6.09 %</td>
<td>0.1723±5.35 %</td>
<td>0.1000±0.66 %</td>
<td>0.0998±0.67 %</td>
<td>0.1167±2.24 %</td>
<td>0.1296±10.1 %</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0533±0.46 %</td>
<td>0.0614±1.84 %</td>
<td>0.0803±0.29 %</td>
<td>0.0781±6.91 %</td>
<td>0.0598±0.45 %</td>
<td>0.0811±4.72 %</td>
<td>0.1109±8.02 %</td>
</tr>
</tbody>
</table>

| RMSE      | Soil            | 0.2949±11.30 % | 0.1922±0.68 % | 0.2205±4.71 % | 0.2097±7.44 % | 0.2316±3.62 % | 0.2612±10.21 % | 0.2435±9.01 % |
|           | Tree            | 0.2912±8.21 %  | 0.1975±1.01 % | 0.2191±1.18 % | 0.2302±4.51 % | 0.2010±0.65 % | 0.2179±9.68 % | 0.2214±6.58 % |
|           | Water           | 0.1001±3.06 %  | 0.3292±0.81 % | 0.3117±2.74 % | 0.3162±4.43 % | 0.3583±2.16 % | 0.3191±5.06 % | 0.2953±5.35 % |
| Mean      | 0.2288±6.55 %   | 0.2396±0.49 %  | 0.2411±2.14 % | 0.2520±2.71 % | 0.2667±1.71 % | 0.2661±5.13 % | 0.2334±3.78 % |

8) Experiment 8 (Running Time): For the purpose of demonstrating the efficiency of SSRDMF and the compared methods, we investigate the average running time of ten times on synthetic data with the different values of $M$, SNR, and ratio & sp, which is listed in Table IV. It is obvious that the uDAS, MLNMF, $L₁/²$-RNMF, $L₁/²$-NMF, and VCA-FCLS algorithms have the lower running time because their structure is relatively simpler. In contrast, despite more running time, the proposed SSRDMF method implements about 100 s and is faster than the $L₁$-DNMF algorithm. In addition, the running time of SSRDMF generally depends on the number of layers and endmembers, and the size of the HSIs.

B. Real Data Experiments

1) AVIRIS Cuprite Dataset: The AVIRIS Cuprite dataset [as shown in Fig. 7(a)] is often used to verify the effectiveness of hyperspectral unmixing methods. It contains 250 × 191 pixels, and each pixel contains 188 bands (i.e., 3–103, 114–147, and 168–220) after removal of noisy bands. The covered wavelength range comprises 0.4–2.5 μm.

Table V lists the corresponding SAD values between each USGS library spectrum and the corresponding endmembers extracted by each method, where 12 minerals are considered. From Table V, we can conclude that all methods achieve satisfactory results for several minerals, such as Andradite WS487, Kaolin/Smeect H89-FR-5 30K, Montmorillonite+Illi CM37, and Sphene HS189.3B. Most importantly, the overall performance of our newly proposed method is the best. Meanwhile, Fig. 8 plots the USGS reference signatures along with the endmembers estimated by SSRDMF on the AVIRIS Cuprite dataset, and Fig. 9 shows the abundance maps estimated by SSRDMF, where darker pixels denote low abundances (and lighter pixels denote high abundances) of the corresponding endmember.

2) Samson Dataset: The Samson dataset is shown in Fig. 7(b). It contains 95 × 95 pixels, and each pixel has 156 bands ranging from 0.401 to 0.889 μm. Generally, three reference endmembers are considered in the subscene: Soil, Tree, and Water.

The performance of each method is obtained by comparing the reference and the estimated ones, shown in Table VI. The proposed SSRDMF outperforms other methods in the estimation of the Water and provides the smallest mean SAD and RMSE. Fig. 10 provides a comparison of the reference signatures versus those estimated by SSRDMF, and Fig. 11 displays the estimated abundance maps by SSRDMF. It can be observed that the estimated endmembers are always in accordance with the ground truth, demonstrating the effectiveness of the proposed method.

3) Jasper Ridge: The Jasper Ridge dataset is shown in Fig. 7(c). It contains 100 × 100 pixels, and each pixel...
remains 198 bands (i.e., 4–107, 113–153, and 167–219) after removal of noisy bands. The covered wavelength range comprises 0.38–2.5 μm. In [1], four reference endmembers are considered in the subscene: Tree, Water, Soil, and Road.

The results are summarized in Table VII and plotted in Figs. 12 and 13. As shown in Table VII, all methods have poor performance when estimating the Road. Nevertheless, SSRDMF can achieve the best results in most cases, especially about the mean SAD and RMSE. From Fig. 12, we can clearly find that each estimated endmember is close to the reference in terms of Tree, Water, and Soil. The abundance maps estimated by SSRDMF are illustrated in Fig. 13.

V. CONCLUSION

In this article, a new algorithm called SSRDMF has been proposed to address the hyperspectral unmixing problem. To obtain better abundances, a deep structure is considered by the way of backpropagation to powerfully encode the observed HSI data. Due to the separation of weights for the encoder and decoder, the number of layers in decoder is set to one, thus matching the extended LMM. To alleviate the interference caused by sparse noise, this structure is extended to RDMF, which not only acquires hidden information but also handles Gaussian noise and sparse noise simultaneously (by introducing a sparse noise term). On this basis, a self-supervised constraint is enforced on the endmember matrix to further improve unmixing performance by adequately exploiting the information contained in spectral signatures. Compared with other state-of-the-art methods, such as L1-DNMF, uDAS, MLNMF, L1/2-RNMF, L1/2-NMF, and VCA-FCLS, the proposed SSRDMF algorithm provides superior unmixing performance using both synthetic and real hyperspectral data.

REFERENCES


Prof. Li is also an Editorial Board Member of the Journal of Southwest Jiaotong University and the Journal of Radars. He was a recipient of the 2018 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society for his service to the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS). He has served as a Guest Editor of Special Issues of the Journal of Real-Time Image Processing, the IEEE JSTARS, and the IEEE Journal on Miniaturization for Air and Space Systems (JMASS), a Program Committee Member of the 26th International Joint Conference on Artificial Intelligence (IJCAI-2107) and the 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MULTITEMP-2019), and the Session Chair of the 2017 International Geoscience and Remote Sensing Symposium (IGARSS-2017) and the 2019 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR-2019). He has been serving as an Associate Editor for the IEEE JSTARS. He is also an active reviewer of over 30 international journals.

Xin-Ru Feng received the B.Sc. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2016, where she is currently pursuing the Ph.D. degree in signal and information processing with the School of Information Science and Technology.

Her research interests include hyperspectral image analysis and processing.

Dong-Hai Zhai (Member, IEEE) received the Ph.D. degree in traffic information engineering and control from Southwest Jiaotong University, Chengdu, China, in 2003.

From 2003 to 2005, he was with the IBM China Research Laboratory, Beijing, China. Since 2006, he has been with the School of Computing and Artificial Intelligence, Southwest Jiaotong University. He has been a Visiting Scholar with Louisiana State University, Baton Rouge, LA, USA, since 2016. His research interests include autonomous driving, digital image processing, computer vision, and pattern recognition.

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

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

Dr. Du is also 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 has served as the Co-Chair of the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was the General Chair of the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held in Shanghai, China, in 2012. She has 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.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was a recipient of the Best Column Award of the IEEE Signal Processing Magazine in 2015 and the Most Highly Cited Paper from 2005 to 2010 in the Journal of Parallel and Distributed Computing. He has served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012 and the President of the Spanish Chapter of the IEEE GRSS from 2012 to 2016. He has served as the Editor-in-Chief of the IEEE Transactions on Geoscience and Remote Sensing (TGRS) from 2013 to 2017 and guest-edited ten special issues on hyperspectral remote sensing for different journals. He is also an Associate Editor of the IEEE Access (also an Outstanding Associate Editor in 2017).