Hyperspectral Unmixing Using Sparsity-Constrained Deep Nonnegative Matrix Factorization With Total Variation

Xin-Ru Feng, Heng-Chao Li, *Senior Member, IEEE*, Jun Li, *Senior Member, IEEE*, Qian Du, *Fellow, IEEE*, Antonio Plaza, *Fellow, IEEE*, and William J. Emery, *Fellow, IEEE*

***Abstract*— Hyperspectral unmixing is an important processing step for many hyperspectral applications, mainly including: 1) estimation of pure spectral signatures (endmembers) and 2) estimation of the abundance of each endmember in each pixel of the image. In recent years, nonnegative matrix factorization (NMF) has been highly attractive for this purpose due to the nonnegativity constraint that is often imposed in the abundance estimation step. However, most of the existing NMF-based methods only consider the information in a single layer while neglecting the hierarchical features with hidden information. To alleviate such limitation, in this paper, we propose a new sparsity-constrained deep NMF with total variation (SDNMF-TV) technique for hyperspectral unmixing. First, by adopting the concept of deep learning, the NMF algorithm is extended to deep NMF model. The proposed model consists of *pretraining stage* and *fine-tuning stage*, where the former pretrains all factors layer by layer and the latter is used to reduce the total reconstruction error. Second, in order to exploit adequately the spectral and spatial information included in the original hyperspectral image, we enforce two constraints on the abundance matrix. Specifically, the $L_{1/2}$ constraint is adopted, since the distribution of each endmember is sparse in the 2-D space. The TV regularizer is further introduced to promote piecewise smoothness in abundance maps. For the optimization of the proposed model, multiplicative update rules are derived using the gradient descent method. The effectiveness and superiority of the SDNMF-TV algorithm are demonstrated by comparing with other unmixing methods on both synthetic and real data sets.**

***Index Terms*— Deep learning (DL), hyperspectral unmixing, nonnegative matrix factorization (NMF), sparsity constraint, total variation (TV).**

Manuscript received October 15, 2017; revised April 9, 2018; accepted May 3, 2018. Date of publication June 5, 2018; date of current version September 25, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61371165 and Grant 61771496 and in part by the Frontier Intersection Basic Research Project for the Central Universities under Grant A0920502051714-5. *(Corresponding author: Heng-Chao Li.)*

X.-R. Feng and H.-C. Li are with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: lihengchao_78@163.com).

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou 510275, China.

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA.

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain.

W. J. Emery is with the Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO 80309 USA.

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

Digital Object Identifier 10.1109/TGRS.2018.2834567

I. Introduction

HYPERSPECTRAL images (HSIs) are acquired by imaging spectrometers in hundreds of contiguous narrow bands [1], [2]. Hence, HSIs exhibit a great potential in acquiring spatial geographic information. A single hyperspectral pixel is inevitably constituted of different materials, due to the low spatial resolution and complicated ground covers, resulting in the concept of "mixed pixels" [3], [4]. The existence of mixed pixels greatly complicates the analysis of HSIs in object detection [5], clustering [6], and classification [7]. Hyperspectral unmixing, whose objective is to decompose a mixed pixel into a collection of constituent materials (called endmembers) and their relative proportions (called abundances), is one of the crucial steps to process HSIs [8].

Mixing models can be divided into two categories: the linear mixing model (LMM) and the nonlinear mixing model (NLMM) [8]. The LMM assumes that the observed pixel spectrum is a linear combination of endmember signatures, weighted by their abundance proportions. Although many studies have illustrated the fact that the NLMM is more accurate on specific surfaces [9]–[11], the LMM is still widely used due to its clear physical significance, simplicity, and efficiency. Based on the LMM, many algorithms have been proposed from geometrical [12], sparse regression [13]–[17], and statistical [18]–[24] perspectives. Specifically, the geometrical class of algorithms is applied to linear unmixing based on the assumption that the vertices of a simplex correspond to the endmembers. Thus, the accuracy of the abundance estimation step is closely related to the endmembers' extraction step. The sparse regression-based algorithms avoid the above dependence. However, they may suffer from incoherent endmember signatures (not derived from the image but often available in a library) and are susceptible to noise and atmospheric interferers. The statistical algorithms estimate simultaneously the endmember matrix and its corresponding abundance matrix by utilizing the statistical properties of the hyperspectral data. Concretely, iterated-constrained endmembers (ICEs) [18], independent component analysis (ICA) [19]–[21], and nonnegative matrix factorization (NMF) [22]–[24] belong to classical statistical algorithms. The ICE is computationally expensive, while the ICA may degrade the unmixing performance due to the unreasonable

hypothesis that the abundances are statistically independent. On the contrary, NMF aims at decomposing the HSIs into two nonnegative matrices simultaneously. In this paper, we focus on the NMF-based methods for linear unmixing purposes.

Due to the fact that nonnegative values in the image matrix have clear physical meaning, NMF has been widely used in spectral unmixing [25]. However, the cost function in this case is nonconvex, which leads to unstable results. To this end, numerous researchers have concentrated on adding some appropriate constraints to improve the performance. The sparsity constraint minimizes the number of materials contained in each mixed pixel. From a mathematical perspective, the L_0 regularizer is a straightforward method and produces the sparsest results. However, since L_0 often leads to an NP-hard problem, it was approximated by the L_1 regularizer in [26] and [27]. Then, owing to the fact that the L_1 regularizer dissatisfies the full additivity constraint, $L_p(0 < p < 1)$ regularizers were considered in [28]–[30] and achieved sparser results. Xu *et al.* [31] illustrated that $L_{1/2}$ was efficient and could be taken as a representative of the $L_p(0 < p < 1)$ regularizers. Qian *et al.* [1] also revealed that $p = 1/2$ was an optimal choice and proposed $L_{1/2}$-NMF by incorporating the $L_{1/2}$ sparsity constraint for unmixing. Many extensions of $L_{1/2}$-NMF have already been reported to alleviate unstable unmixing results by introducing sparse noise [4], structural information [32], and region-based structure [33].

In addition, researchers worked on improving unmixing performance by exploiting the spatial-contextual information contained in the hyperspectral data [34]–[36]. For example, Lu *et al.* [34] introduced an intrinsic manifold structure to sparsity-constrained NMF and proposed the graph-regularized $L_{1/2}$-NMF method. By integrating the geometric structure in HSIs, Yang *et al.* [35] proposed the geometric NMF method and obtained a more desirable result. In [36], a total variation (TV) regularized reweighed sparse NMF method was proposed by simultaneously considering the spatial information and the sparsity constraint. The TV model, proposed by Rudin *et al.* [37], was shown to effectively promote piecewise smoothness in the estimation of abundance maps.

Despite the commendable success of the above methods in spectral unmixing, there is still room to further improve the unmixing performance. For instance, the NMF-based methods with two factors only consider the information in a single layer, ignoring the fact that the original data contain hierarchical features with hidden information. Recently, a neural network (NN) is recognized an effective technique for hyperspectral unmixing [38]–[43]. However, the NN-based methods are also restricted to a shallow structure, resulting in gradient diffusion and failure to extract hidden features. To solve these issues, Hinton *et al.* [44] proposed the deep leaning (DL) strategy for extending the NN-based methods to deep structure, which has been applied in HSIs' processing and face recognition [45]–[48]. Inspired by the success of DL, we adopt its concept into the NMF-based methods and propose a novel sparsity-constrained deep NMF with TV (SDNMF-TV) algorithm, which simultaneously explores the spectral sparsity as well as the spatial smoothness of fractional abundances. The main contributions of this paper could be summarized as follows.

1) Through adopting the idea of DL, we first extend the NMF algorithm to deep NMF model for exploring hierarchical features in spectral unmixing. The unmixing process based on the deep NMF structure consists of two stages, where the first stage pretrains all factors layer by layer and the second stage fine-tunes all factors until a stopping criterion is reached.
2) Since the number of endmembers associated with each mixed pixel is generally much lower than the total number of endmembers in a spectral library, the requirement of sparsity in an abundance matrix must be satisfied. Accordingly, the $L_{1/2}$ regularizer is introduced into the deep NMF model to propose sparsity-constrained deep NMF (SDNMF) for hyperspectral unmixing.
3) We further integrate the TV regularizer into SDNMF and propose the SDNMF-TV model for hyperspectral unmixing, which could promote piecewise smoothness and reduce the negative effects of noise in the abundance maps.
4) For the estimation of endmember signatures and their corresponding abundances, it is a crucial issue to optimize effectively the proposed model. Specifically, the multiplicative update rules are derived for these two stages using the gradient descent (GD) method.

The remainder of this paper is organized as follows. Section II introduces the related work. Section III describes the proposed deep NMF framework in detail. The proposed SDNMF and SDNMF-TV are presented in Section IV. In Section V, experiments and related analysis using synthetic and real hyperspectral data sets are provided. Finally, Section VI concludes this paper with some remarks and hints at plausible future research.

II. Related Work

A. Linear Mixing Model

The LMM assumes that an observed pixel spectrum in an HSI can be produced by a linear combination of endmember signatures and their corresponding abundances. The matrix formulation of the LMM can be described as

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

where $\mathbf{X} \in \mathbb{R}^{B\times P}$ refers to the observation matrix, with B bands and P pixels, $\mathbf{A} \in \mathbb{R}^{B\times M}$ consists of the spectral signatures with M endmembers, $\mathbf{S} \in \mathbb{R}^{M\times P}$ represents the abundance matrix for all endmembers, and $\mathbf{N} \in \mathbb{R}^{B\times P}$ represents the noise matrix. In general, two constraints are imposed on $\mathbf{S}$, i.e., the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC), which can be formulated as

$$S_{i,j} \geq 0, \quad \sum_i S_{i,j} = 1 \tag{2}$$

where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, P$. These two constraints are supported by the physical meaning that all proportions of the materials in a mixed pixel are nonnegative and their sum is one.

B. Nonnegative Matrix Factorization and Its Variants

In recent years, NMF and its variants have achieved the state-of-the-art performance in spectral unmixing of HSIs. The cost function of the NMF-based methods can be presented as

$$\mathcal{C}_{\text{NMF}} = \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda\|\mathbf{S}\|_q \tag{3}$$

where the operators $\|\cdot\|_F$ and $\|\cdot\|_q$ denote the Frobenius norm and L_q regularizer, respectively, and λ is used to regulate the effect of the sparsity constraint. Specifically, the different unmixing methods depend on the values of λ and q. For instance, the following holds.

1) The problem of (3) with $\lambda = 0$ is the classical NMF [49].
2) By setting $\lambda \neq 0$ and $q = 1$, (3) is turned into L_1-NMF [26], [27], where $\|\mathbf{S}\|_1 = \sum_{i,j=1}^{M,P} |s_{i,j}|$ is utilized to achieve sparse results.
3) Similarly, we regard (3) with $\lambda \neq 0$ and $q = 1/2$ as $L_{1/2}$-NMF [1], [31], which could be a representative of the $L_p (0 < p < 1)$ regularizers.

Notice that the ANC constraint is certainly satisfied by the NMF-based methods. Meanwhile, the ASC constraint is enforced by redefining the observation and spectral signature matrix as [32]

$$\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \delta\mathbf{1} \end{bmatrix}, \quad \bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \delta\mathbf{1} \end{bmatrix} \tag{4}$$

where δ controls the impact of the ASC constraint and $\mathbf{1}$ is a row vector with all elements equal to one. Then, the multiplicative iterative rules can be summarized as [1], [49]

$$\mathbf{A} = \mathbf{A}. * (\mathbf{XS}^T)./(\mathbf{ASS}^T) \tag{5a}$$

$$\mathbf{S} = \mathbf{S}. * (\bar{\mathbf{A}}^T\bar{\mathbf{X}})./(\bar{\mathbf{A}}^T\bar{\mathbf{A}}\mathbf{S} + \lambda q\mathbf{S}^{q-1}) \tag{5b}$$

where $.*$ and $./$ stand for the elementwise multiplication and division, respectively, $(\cdot)^T$ denotes the transpose operation to a matrix, and $\mathbf{S}^{q-1}$ is calculated by

$$\mathbf{S}^{q-1} = \sum_{i,j=1}^{M,P} (s_{i,j})^{q-1}. \tag{6}$$

III. Proposed Deep Nonnegative Matrix Factorization Model

In this section, we first describe the newly proposed deep NMF model. Then, effective multiplicative iterative algorithms consisting of two stages are deduced in detail to estimate simultaneously endmember signatures and abundances.

A. Formulation of Proposed Deep NMF Model

The basic NMF (and its variants) can be regarded as a single-layer learning process, illustrated in Fig. 1(a), which learns simultaneously the endmember signatures $\mathbf{A}$ and the abundance matrix $\mathbf{S}$ from the observation matrix $\mathbf{X}$. By stacking one hidden layer on the bottom of each layer, i.e., the abundance matrix $\mathbf{S}_1$ acquired over the single layer could be factorized into $\mathbf{A}_2$ and $\mathbf{S}_2$. As such, we extend the shallow NMF into a two-layer NMF structure, as shown in Fig. 1(b).

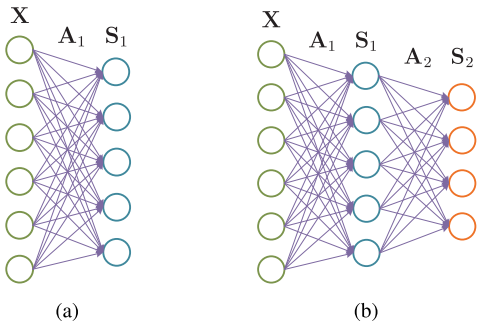


Fig. 1. Schematic diagram for (a) one-layer and (b) two-layer learning.

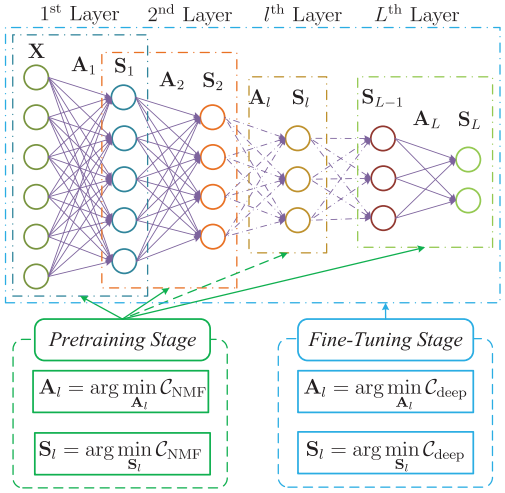


Fig. 2. Schematic diagram of the proposed deep NMF model.

Fig. 2 shows the framework of the deep NMF model, which factorizes the observation matrix $\mathbf{X}$ into $L+1$ factors, given by

$$\mathbf{X} = \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_L\mathbf{S}_L \tag{7}$$

where $\mathbf{A}_l \in \mathbb{R}^{M_{l-1}\times M_l}$, $l = 1, 2, \ldots, L$, and M_0 is used to represent the number of bands B.

This means that the observation matrix could be decomposed in L different ways, where the abundance matrix $\mathbf{S}$ of different layers is given by

$$\begin{aligned} \mathbf{S}_{L-1} &= \mathbf{A}_L\mathbf{S}_L \\ &\vdots \\ \mathbf{S}_2 &= \mathbf{A}_3 \ldots \mathbf{A}_L\mathbf{S}_L \\ \mathbf{S}_1 &= \mathbf{A}_2 \ldots \mathbf{A}_L\mathbf{S}_L. \end{aligned} \tag{8}$$

These hidden representations $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$ are further restricted to be nonnegative as well. As shown in Fig. 2, the deep NMF model is composed of two stages.

1) *Pretraining Stage:* This stage aims at pretraining each layer by minimizing the cost function in (3) with $\lambda = 0$. Specifically, in the first layer, the observation matrix $\mathbf{X}$ is decomposed into $\mathbf{A}_1$ and $\mathbf{S}_1$. Then, in the second layer, $\mathbf{S}_1$ is decomposed into $\mathbf{A}_2$ and $\mathbf{S}_2$. The same process is continued until the maximum number of layers (L) is reached.

2) *Fine-Tuning Stage:* To decrease the total reconstruction error, we fine-tune all factors acquired at the previous stage according to the following cost function:

$$\mathcal{C}_{\text{deep}} = \frac{1}{2}\|\mathbf{X} - \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_L\mathbf{S}_L\|_F^2. \tag{9}$$

Finally, once we obtain all optimal factors, $\mathbf{A}$ and $\mathbf{S}$ are, respectively, calculated by

$$\mathbf{A} = \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_L \tag{10a}$$
$$\mathbf{S} = \mathbf{S}_L. \tag{10b}$$

B. Optimization of Deep NMF

1) Update Rules for Pretraining Stage: We approximate the factors layerwise greedily using the classical NMF algorithm for all layers [49], i.e.,

$$\mathbf{A} = \mathbf{A}. * (\mathbf{X}\mathbf{S}^T)./(\mathbf{A}\mathbf{S}\mathbf{S}^T) \tag{11a}$$
$$\mathbf{S} = \mathbf{S}. * (\bar{\mathbf{A}}^T\bar{\mathbf{X}})./(\bar{\mathbf{A}}^T\bar{\mathbf{A}}\mathbf{S}). \tag{11b}$$

2) Update Rules for Fine-Tuning Stage: On this stage, all the factors obtained at the former stage are fine-tuned. To determine the update rules, we rewrite the cost function given in (9) as a detailed form

$$\mathcal{C}_{\text{deep}} = \frac{1}{2}\|\mathbf{X} - \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_{l-1}\mathbf{A}_l\mathbf{A}_{l+1} \ldots \mathbf{A}_L\mathbf{S}_L\|_F^2 \tag{12}$$

where we have defined Ψ_{l-1} and $\widetilde{\mathbf{S}}_l$ as

$$\Psi_{l-1} = \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_{l-1} \tag{13a}$$
$$\widetilde{\mathbf{S}}_l = \mathbf{A}_{l+1} \ldots \mathbf{A}_L\mathbf{S}_L \tag{13b}$$

with $\widetilde{\mathbf{S}}_l$ being the reconstruction of the lth layer's abundance matrix. It should be noted that $\Psi_{l-1} = \mathbf{I}$ if and only if $l = 1$, here $\mathbf{I}$ represents an identity matrix. Hence, the cost function given in (12) becomes

$$\mathcal{C}_{\text{deep}} = \frac{1}{2}\|\mathbf{X} - \Psi_{l-1}\mathbf{A}_l\widetilde{\mathbf{S}}_l\|_F^2. \tag{14}$$

The local optimal $\mathbf{A}_l$ and $\mathbf{S}_l$ can be obtained by minimizing the above function. In this paper, to estimate the optimal $\mathbf{A}_l$ and $\mathbf{S}_l$, the multiplicative update rules are calculated by using the GD method. First, the gradients are solved by differentiating (14) with respect to $\mathbf{A}_l$ and $\mathbf{S}_l$, respectively, as

$$\nabla_{\mathbf{A}_l}\mathcal{C}_{\text{deep}} = -\Psi_{l-1}^{\mathrm{T}}\mathbf{X}\widetilde{\mathbf{S}}_l^T + \Psi_{l-1}^T\Psi_{l-1}\mathbf{A}_l\widetilde{\mathbf{S}}_l\widetilde{\mathbf{S}}_l^T \tag{15a}$$
$$\nabla_{\mathbf{S}_l}\mathcal{C}_{\text{deep}} = -\mathbf{A}_l^{\mathrm{T}}\Psi_{l-1}^T\mathbf{X} + \mathbf{A}_l^{\mathrm{T}}\Psi_{l-1}^T\Psi_{l-1}\mathbf{A}_l\mathbf{S}_l. \tag{15b}$$

After selecting two appropriate step sizes, the update rules for the lth layer on the *fine-tuning stage* are obtained as

$$\mathbf{A}_l = \mathbf{A}_l. * \left(\Psi_{l-1}^T\mathbf{X}\widetilde{\mathbf{S}}_l^T\right)./\left(\Psi_{l-1}^T\Psi_{l-1}\mathbf{A}_l\widetilde{\mathbf{S}}_l\widetilde{\mathbf{S}}_l^T\right) \tag{16a}$$
$$\mathbf{S}_l = \mathbf{S}_l. * \left(\bar{\Psi}_l^T\bar{\mathbf{X}}\right)./\left(\bar{\Psi}_l^T\bar{\Psi}_l\mathbf{S}_l\right) \tag{16b}$$

where the ASC is satisfied for $\mathbf{S}_l$ by augmenting the $\Psi_{l-1}\mathbf{A}_l$ as

$$\bar{\Psi}_l = \begin{bmatrix}\Psi_{l-1}\mathbf{A}_l \\ \delta\mathbf{1}\end{bmatrix}. \tag{17}$$

The proposed deep NMF algorithm is summarized in Algorithm 1. Its implementation details are described in Section IV-C.

Algorithm 1 Proposed Deep NMF Algorithm

Input: Observation matrix $\mathbf{X}$;
Number of endmembers M;
Parameters L, δ, and $T_{\max}$.
Pretraining stage:
Repeat for l
Initialize: $\mathbf{A}_l$ and $\mathbf{S}_l$ by the VCA and FCLS methods.
Repeat
Update $\mathbf{A}_l$ by (11a);
Update $\mathbf{S}_l$ by (11b).
until stopping criterion is satisfied.
until L is reached.
Fine-tuning stage:
Repeat
for all layers **do**
Update $\mathbf{A}_l$ by (16a);
Update $\mathbf{S}_l$ by (16b).
end for
until stopping criterion is satisfied.
Output: The endmember signature matrix $\mathbf{A}$ by (10a);
the abundance matrix $\mathbf{S}$ by (10b).

C. Computational Complexity Analysis

According to Algorithm 1, the main time cost is spent on updating rules giving in (11a), (11b), (16a), and (16b). The computational complexity for the *pretraining stage* is $\mathcal{O}(\sum_{l=1}^{L} t_l(PM_{l-1}M_l + PM_l^2 + M_{l-1}M_l^2))$, where t_l is the number of iterations; for the *fine-tuning stage*, it is $\mathcal{O}(t_f(BPM_1 + BM_1^2 + PM_1^2 + \sum_{l=2}^{L}(BPM_l + BM_l^2 + PM_l^2 + PM_{l-1}M_l + M_{l-1}^2M_l + M_{l-1}M_l^2))$, in which t_f is the number of iterations.

IV. Proposed Sparsity-Constrained Deep NMF With Total Variation

A. SDNMF-TV Model

The cost function of NMF is nonconvex, leading to unstable results. Hence, applying appropriate constraints is an alternative scheme to improve unmixing performance.

In real images, the number of materials contained in each mixed pixel is generally much lower than the total number of materials in the spectral library. Therefore, it is reasonable to add the sparsity constraint to the cost functions. For the methods that use L_p regularizer, Qian *et al.* [1] have already demonstrated that the $L_{1/2}$-NMF is better than the other L_q-NMF algorithms by measuring sparsity and computational complexity. To this end, the cost function of SDNMF with the $L_{1/2}$ regularizer is given by

$$\mathcal{C} = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_{1/2} \tag{18a}$$
$$\mathcal{C}_{\text{deep}} = \frac{1}{2}\|\mathbf{X} - \Psi_{l-1}\mathbf{A}_l\widetilde{\mathbf{S}}_l\|_F^2 + \lambda\|\mathbf{S}_l\|_{1/2} \tag{18b}$$

where (18a) and (18b) are related to the *pretraining stage* and *fine-tuning stage*, respectively.

In addition to spectral information, spatial information is another important aspect of HSIs. The TV regularization is

further integrated into the SDNMF model to enforce the derivation of piecewise smoothness abundance maps. The TV norm of an HSI is defined as [36]

$$\|\mathbf{S}\|_{\mathrm{HTV}} = \sum_{j=1}^{M} \|\mathcal{F}\mathbf{S}^j\|_{\mathrm{TV}} \tag{19}$$

where $\mathbf{S}^j$ denotes the jth row of the matrix $\mathbf{S}$ and $\mathcal{F}$ represents the reshape operation from a vector with P pixels to a matrix with $m \times n$. Notice that the values of m and n correspond to the geometric dimensions of an HSI cube. The TV model proposed by Rudin *et al.* [37] can promote piecewise smoothness effectively. For an image $\mathbf{y}$ with $m \times n$ elements, the anisotropic TV norm is reformed as

$$\begin{aligned}\|\mathbf{y}\|_{\mathrm{TV}} = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}\{|y_{i,j} - y_{i+1,j}| + |y_{i,j} - y_{i,j+1}|\} \\ &+ \sum_{i=1}^{m-1}|y_{i,n} - y_{i+1,n}| + \sum_{j=1}^{n-1}|y_{m,j} - y_{m,j+1}|. \end{aligned} \tag{20}$$

Combining the sparsity constraint with spatial information, we propose a new SDNMF-TV model, whose cost function can be represented as

$$\mathcal{C} = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_{1/2} + \alpha\|\mathbf{S}\|_{\mathrm{HTV}} \tag{21a}$$

$$\mathcal{C}_{\mathrm{deep}} = \frac{1}{2}\|\mathbf{X} - \Psi_{l-1}\mathbf{A}_l\widetilde{\mathbf{S}}_l\|_F^2 + \lambda\|\mathbf{S}_l\|_{1/2} + \alpha\|\mathbf{S}_l\|_{\mathrm{HTV}} \tag{21b}$$

where α adjusts the strength of the piecewise smoothness. The SDNMF-TV method becomes SDNMF when α is set as 0, and it is the deep NMF model when λ is futher set as 0. Analogously, (21a) and (21b) are given for the *pretraining stage* and *fine-tuning stage*, respectively.

B. Optimization of the SDNMF-TV Model

Due to incorporation of the sparsity constraint and TV regularizer, the optimization method of SDNMF-TV is slightly different from that of the deep NMF model. More details about the derivation procedure are described as follows.

1) Update Rules for Pretraining Stage: To effectively deduce the multiplicative update rules, an auxiliary matrix $\mathbf{L}$ is introduced to (21a) and the cost function is transformed as

$$\mathcal{C} = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_{1/2} + \alpha\|\mathbf{L}\|_{\mathrm{HTV}} \quad \text{s.t. } \mathbf{L} = \mathbf{S}. \tag{22}$$

The matrix $\mathbf{S}$ can be regarded as the combination of $\mathbf{L}$ with noise. For convenience of calculation, the constraint $\mathbf{L} = \mathbf{S}$ can be absorbed into the cost function [36], i.e.,

$$\mathcal{C} = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_{1/2} + \frac{\mu}{2}\|\mathbf{L} - \mathbf{S}\|_F^2 + \alpha\|\mathbf{L}\|_{\mathrm{HTV}} \tag{23}$$

where μ is a parameter to control the similarity between $\mathbf{S}$ and $\mathbf{L}$, and a larger value of μ enforces higher similarity of the two matrices.

By taking into account the fact that the sparsity constraint and the TV regularizer are applied to the abundance matrix S, the update rule of $\mathbf{A}$ is the same as (11a). For the solution of $\mathbf{S}$, we first differentiate (23) with respect to $\mathbf{S}$, i.e.,

$$\nabla_{\mathbf{S}}\mathcal{C} = \mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{S} - \mathbf{A}^{\mathrm{T}}\mathbf{X} + \frac{\lambda}{2}\mathbf{S}^{-1/2} + \mu(\mathbf{S} - \mathbf{L}). \tag{24}$$

After selecting a reasonable step size, we can obtain the update rule for $\mathbf{S}$ as

$$\mathbf{S} = \mathbf{S}. * (\bar{\mathbf{A}}^T\bar{\mathbf{X}} + \mu\mathbf{L}). \Big/ \left(\bar{\mathbf{A}}^T\bar{\mathbf{A}}\mathbf{S} + \frac{\lambda}{2}\mathbf{S}^{-1/2} + \mu\mathbf{S}\right). \tag{25}$$

The method for deriving the update rule of $\mathbf{L}$ is distinct from that of $\mathbf{A}$ and $\mathbf{S}$. Specifically, the update rule of $\mathbf{L}$ is obtained by minimizing the following function:

$$\frac{\mu}{2}\|\mathbf{L} - \mathbf{S}\|_F^2 + \alpha\|\mathbf{L}\|_{\mathrm{HTV}}. \tag{26}$$

By combining (19) and (26), the following optimization function is formulated as:

$$\mathbf{L} = \arg\min_{\mathbf{L}} \sum_{j=1}^{M}\left(\frac{\mu}{2}\|\mathcal{F}\mathbf{L}^j - \mathcal{F}\mathbf{S}^j\|_F^2 + \alpha\|\mathcal{F}\mathbf{L}^j\|_{\mathrm{TV}}\right) \tag{27}$$

where the fast gradient projection algorithm [50] is applied to solve (27) for each j.

2) Update Rules for Fine-Tuning Stage: Referring to the fine-tuning process in the deep NMF model and the pretraining one in the SDNMF-TV model, the update rules for the *fine-tuning stage* are given as

$$\mathbf{A}_l = \mathbf{A}_l. * \left(\Psi_{l-1}^T\mathbf{X}\widetilde{\mathbf{S}}_l^T\right). \Big/ \left(\Psi_{l-1}^T\Psi_{l-1}\mathbf{A}_l\widetilde{\mathbf{S}}_l\widetilde{\mathbf{S}}_l^T\right) \tag{28a}$$

$$\mathbf{S}_l = \mathbf{S}_l. * \left(\bar{\Psi}_l^T\bar{\mathbf{X}} + \mu\mathbf{L}_l\right). \Big/ \left(\bar{\Psi}_l^T\bar{\Psi}_l\mathbf{S}_l + \frac{\lambda}{2}\mathbf{S}_l^{-1/2} + \mu\mathbf{S}_l\right) \tag{28b}$$

$$\mathbf{L}_l = \min_{\mathbf{L}_l}\sum_{j=1}^{M}\left(\frac{\mu}{2}\left\|\mathcal{F}\mathbf{L}_l^j - \mathcal{F}\mathbf{S}_l^j\right\|_F^2 + \alpha\left\|\mathcal{F}\mathbf{L}_l^j\right\|_{\mathrm{TV}}\right). \tag{28c}$$

The newly proposed SDNMF-TV algorithm is summarized in Algorithm 2.

C. Implementation Issues

Here, several issues will be discussed to implement the proposed algorithm. The first issue is how to determine the number of endmembers, for which the studies have reported many effective methods such as virtual dimensionality [51] and HySime [52]. However, it is another independent topic. Thus, we set the number of endmembers manually. Specifically, on synthetic data, the number of endmembers is given according to the generation process. On real data, it is provided following several previous studies.

The second issue involves the initialization of $\mathbf{A}_l$ and $\mathbf{S}_l$. In our implementation, the matrix $\mathbf{A}_l$ is initialized by using the vertex component analysis (VCA) algorithm [12], and the matrix $\mathbf{S}_l$ is initialized by using the fully constrained least squares (FCLS) algorithm [53].

For the $L_{1/2}$ regularizer, not all the elements in $\mathbf{S}$ (or $\mathbf{S}_l$) are updated following (25) [or (28b)]. As shown in [4], we set a predefined threshold (1×10^{-4}). For those elements less than the threshold, we omit the additional term corresponding to the $L_{1/2}$ sparsity operator.

Algorithm 2 SDNMF-TV for Spectral Unmixing

Input: Observation matrix $\mathbf{X}$;
Number of endmembers M;
Parameters L, δ, $T_{\max}$, λ, α, and μ.
Pretraining stage:
Repeat for l
Initialize: $\mathbf{A}_l$ and $\mathbf{S}_l$ by the VCA and FCLS methods.
Repeat
Update $\mathbf{A}_l$ by (11a);
Update $\mathbf{S}_l$ by (25);
Update $\mathbf{L}_l$ by (27).
until stopping criterion is satisfied.
until L is reached.
Fine-tuning stage:
Repeat
for all layers **do**
Update $\mathbf{A}_l$ by (28a);
Update $\mathbf{S}_l$ by (28b);
Update $\mathbf{L}_l$ by (28c).
end for
until stopping criterion is satisfied.
Output: The endmember signature matrix $\mathbf{A}$ by (10a);
the abundance matrix $\mathbf{S}$ by (10b).

The last issue is about the stopping criteria. First of all, we define the reconstruction error $\mathcal{J}$ as

$$\mathcal{J} = \|\mathbf{S}_{l-1} - \mathbf{A}_l\mathbf{S}_l\|_F^2 \tag{29a}$$

$$\mathcal{J} = \|\mathbf{X} - \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_L\mathbf{S}_L\|_F^2 \tag{29b}$$

where (29a) and (29b) are utilized for the *pretraining stage* and *fine-tuning stage*, respectively. If the number of iterations reaches a preset maximum number ($T_{\max}$) or the stopping criterion ($|\mathcal{J}^{(t-1)} - \mathcal{J}^{(t)}|/\mathcal{J}^{(t-1)} \leq \epsilon$) is met, the proposed algorithm will be stopped, in which t represents the number of iterations. ϵ is the error threshold, which is set to be 1×10^{-3} in the experiments. It is worth mentioning that $\mathbf{X}$ is equivalent to $\mathbf{S}_{l-1}$ if and only if $l = 1$ in (29a).

V. Experimental Results and Discussion

In order to demonstrate the effectiveness of our newly proposed method, experiments on synthetic and real hyperspectral data have been implemented. Synthetic data are utilized, because its true endmembers and corresponding abundances are completely known. The real hyperspectral data sets used in our experiments are the Samson and Cuprite data sets, respectively. The results of MLNMF [54], $L_{1/2}$-NMF [1], and VCA-FCLS are used for comparative purposes. For the proposed method, the maximum number of iterations $T_{\max}$ is selected as 500, the parameters δ and μ are fixed to 15 and 1×10^3, respectively, based on [36] and [32].

For quantitative assessments, the spectral angle distance (SAD) is adopted to measure the difference between the mth original endmember ($\mathbf{A}_m$) and its estimate ($\hat{\mathbf{A}}_m$), which is defined as

$$\mathrm{SAD}_m = \arccos\left(\frac{\mathbf{A}_m^T\hat{\mathbf{A}}_m}{\|\mathbf{A}_m^T\|\|\hat{\mathbf{A}}_m\|}\right). \tag{30}$$

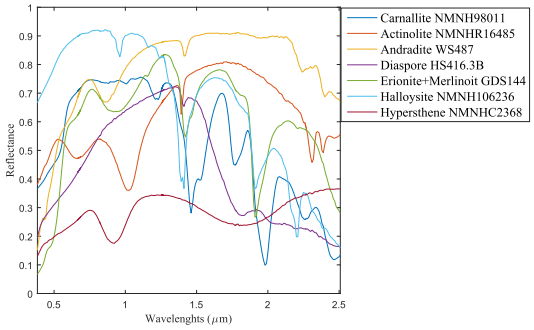


Fig. 3. Spectral signatures of seven endmembers chosen from the USGS library in our experiments.

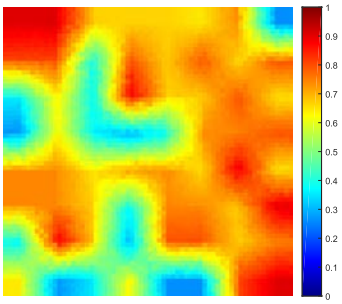


Fig. 4. Synthetic image at band 1.

Similarly, the abundance angle distance (AAD) can be used to measure the difference between the mth original ($\mathbf{S}_m$) and the estimated abundance ($\hat{\mathbf{S}}_m$), which is defined as

$$\mathrm{AAD}_m = \arccos\left(\frac{\mathbf{S}_m\hat{\mathbf{S}}_m^T}{\|\mathbf{S}_m\|\|\hat{\mathbf{S}}_m^T\|}\right). \tag{31}$$

A. Synthetic Data Experiments

Synthetic data are generated using the U.S. Geological Survey (USGS) spectral library, which contains nearly 500 typical minerals and is available online at http://speclab.cr.usgs.gov/spectral.lib06. It comprises spectral signatures with reflectance values given in 224 spectral bands and distributed uniformly over the interval 0.4–2.5 μm. In this paper, the endmember matrix $\mathbf{A}$ is constructed by seven spectral signatures chosen from the USGS randomly, as shown in Fig. 3. Their abundances are generated following the methodology provided in [55]: 1) a scene with a size of $z^2 \times z^2$ ($z = 8$) is divided into $z \times z$ regions and each region is filled up by one of the above endmembers; 2) mixed pixels are produced by a spatial low-pass filter ($(z + 1) \times (z + 1)$); and 3) to further remove pure pixels, the pixels with abundances being larger than θ ($\theta = 0.8$) are replaced by a mixture composed of all endmembers with equal weights, and afterward, the distributions of seven endmembers in the scene are obtained and stored in $\mathbf{S}$; and 4) the synthetic data, shown in Fig. 4, is generated according to $\mathbf{X} = \mathbf{AS}$.

Experiment 1 (Parameter Analysis): For the proposed method, parameters λ and α regulate the impact of sparsity constraint and piecewise smoothness, respectively. In this

TABLE I

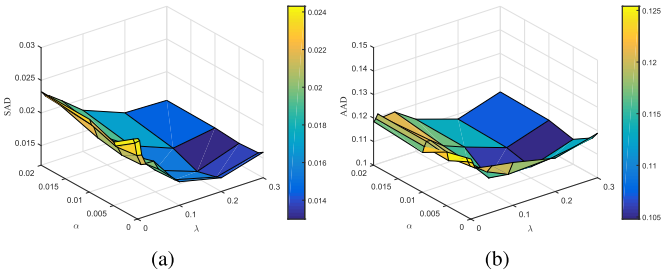
COMPARISON OF THE SAD ON THE SYNTHETIC DATA FOR DIFFERENT METHODS IN TERMS OF NOISE. WE LABEL THE BEST RESULTS IN BOLD TYPEFACE AT EACH DESIGNATED SNR VALUE, AND THE SECOND BEST RESULTS ARE UNDERLINED

Noise	SNR (dB)	SDNMF-TV	SDNMF	MLNMF	$L_{1/2}$-NMF	VCA-FCLS
Gaussian noise	10	**0.0540**	0.0592	0.0737	0.0721	0.0873
	15	**0.0227**	0.0236	0.0349	0.0374	0.0477
	20	**0.0130**	0.0141	0.0218	0.0229	0.0327
	25	**0.0106**	0.0107	0.0144	0.0147	0.0179
	30	**0.0072**	0.0080	0.0105	0.0110	0.0119
Correlated noise	30	**0.0071**	0.0078	0.0104	0.0096	0.0113

TABLE II

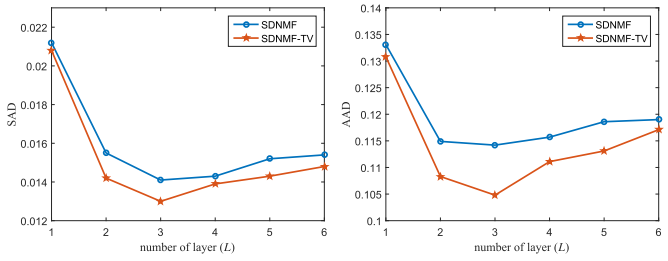
COMPARISON OF THE AAD ON THE SYNTHETIC DATA FOR DIFFERENT METHODS IN TERMS OF NOISE. WE LABEL THE BEST RESULTS IN BOLD TYPEFACE AT EACH DESIGNATED SNR VALUE, AND THE SECOND BEST RESULTS ARE UNDERLINED

Noise	SNR (dB)	SDNMF-TV	SDNMF	MLNMF	$L_{1/2}$-NMF	VCA-FCLS
Gaussian noise	10	**0.3084**	0.3189	0.3340	0.3479	0.3500
	15	**0.1737**	0.1753	0.1953	0.2116	0.2202
	20	**0.1048**	0.1142	0.1194	0.1318	0.1693
	25	**0.0881**	0.1017	0.0885	0.1179	0.1215
	30	**0.0554**	0.0636	0.0644	0.0962	0.1027
Correlated noise	30	**0.0549**	0.0630	0.0637	0.0874	0.0918



(a) (b)

Fig. 5. Performance of SDNMF ($\alpha = 0$) and SDNMF-TV with respect to parameters λ and α in terms of (a) SAD and (b) AAD.



(a) (b)

Fig. 6. Comparison of (a) SAD and (b) AAD for SDNMF and SDNMF-TV versus the number of layers L.

experiment, we investigate the performance of our algorithm in terms of λ and α when the signal-to-noise ratio (SNR) is set to 20 dB. To reduce the computational burden, λ is defined in a finite set $\{0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3\}$, and α is defined in a finite set $\{0, 0.001, 0.005, 0.01, 0.02\}$. Fig. 5(a) shows the SAD achieved by SDNMF and SDNMF-TV when these two parameters λ and α vary under $L = 3$, and Fig. 5(b) gives the AAD results. As can be seen, when the value of parameter λ is less than 0.2, both the SAD and AAD values will increase gradually, while the value of λ exceeds 0.2 and the values of SAD and AAD increase as well. Similarly, the values of the two metrics will increase slowly as α exceeds or is less than 0.005. Namely, we could observe that the best values of λ and α are in the vicinity of 0.2 and 0.005, respectively. The value of λ (i.e., 0.2) is larger than that of α (i.e., 0.005), revealing that the sparsity constraint has a predominant impact on the final result.

Experiment 2 (Analysis to the Number of Layers): Since the core characteristics of a deep NMF method and its extensions are to factorize the observation matrix into many factors, the number of layers is a key parameter. In this experiment, we assign the number of layers L from 1 to 6 with SNR = 20 dB.

Fig. 6 presents a comparison of the obtained results in terms of SAD and AAD. We can observe clearly that the curves plotted by these two metrics for SDNMF and SDNMF-TV exhibit a similar trend. Specifically, the values of the SAD and AAD when $L = 1$ are larger than others, verifying that the deep NMF model is better than the shallow one. The performance of our proposed method is improved gradually with the value of L being increased, while it is degraded when L increases to a certain level. For the synthetic data, these two metrics reach their minima at $L = 3$; in other words, the optimal performances of SDNMF and SDNMF-TV are achieved by using three-layer structure. Thus, we set L as 3 to extract endmember signatures and retrieve the corresponding abundances in the following experiments.

Experiment 3 (Robustness Analysis): In this experiment, to verify the superiority and robustness of the proposed algorithm when data are contaminated by noise, the synthetic data are degraded by Gaussian noise and correlated noise. Specifically, the Gaussian noise contains five levels of SNR, i.e., 10, 15, 20, 25, and 30 dB. Meanwhile, the correlated noise only has one level of SNR, i.e., SNR = 30 dB, which is generated by low-pass filtering independent and identically

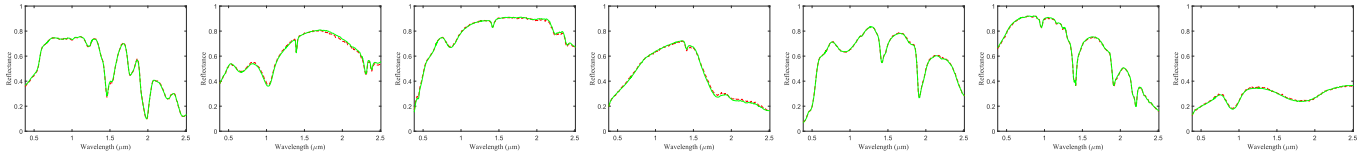


Fig. 7. Comparison of the reference spectra (green solid line) with the endmember signatures estimated by SDNMF-TV (red dashed line) on the synthetic data. (a) Carnallite NMNH98011. (b) Actinolite NMNHR16485. (c) Andradite WS487. (d) Diaspore HS416.3B. (e) Erionite+Merlinoit GDS144. (f) Halloysite NMNH106236. (g) Hypersthene NMNHC2368.

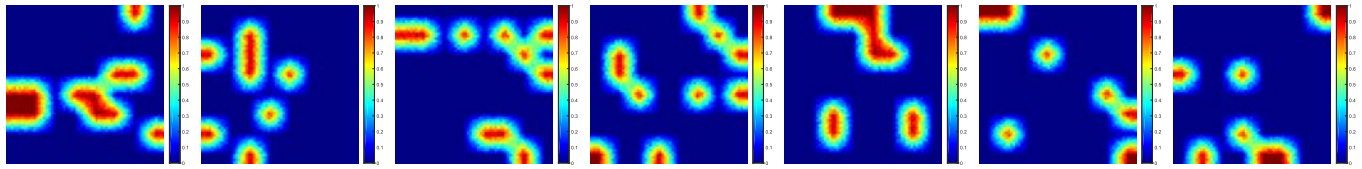


Fig. 8. True fractional abundance maps for seven endmembers. (a) Carnallite NMNH98011. (b) Actinolite NMNHR16485. (c) Andradite WS487. (d) Diaspore HS416.3B. (e) Erionite + Merlinoit GDS144. (f) Halloysite NMNH106236. (g) Hypersthene NMNHC2368.

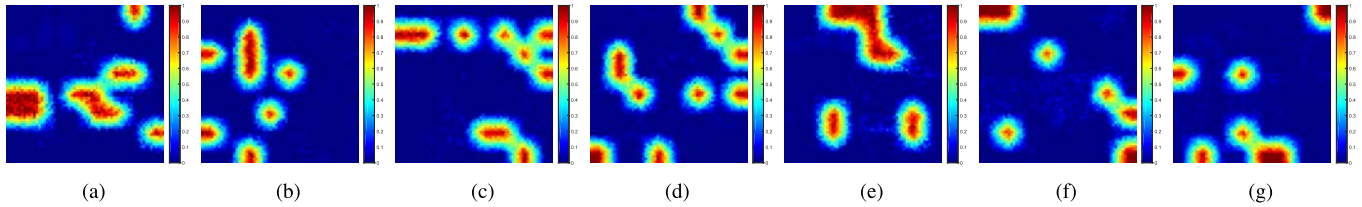


Fig. 9. Fractional abundance maps estimated by SDNMF-TV for seven endmembers. (a) Carnallite NMNH98011. (b) Actinolite NMNHR16485. (c) Andradite WS487. (d) Diaspore HS416.3B. (e) Erionite + Merlinoit GDS144. (f) Halloysite NMNH106236. (g) Hypersthene NMNHC2368.

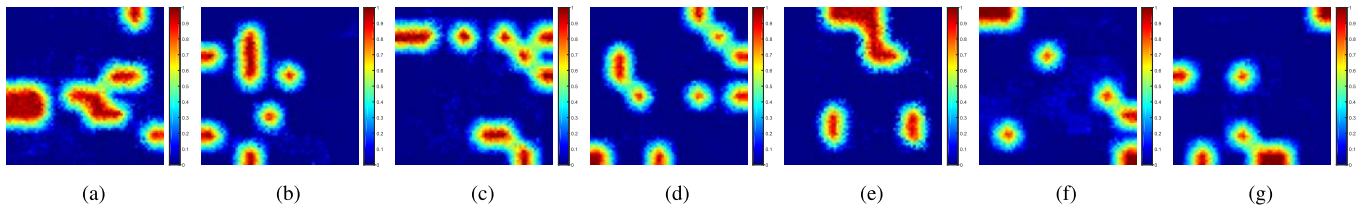


Fig. 10. Fractional abundance maps estimated by SDNMF for seven endmembers. (a) Carnallite NMNH98011. (b) Actinolite NMNHR16485. (c) Andradite WS487. (d) Diaspore HS416.3B. (e) Erionite + Merlinoit GDS144. (f) Halloysite NMNH106236. (g) Hypersthene NMNHC2368.

distributed Gaussian noise with a normalized cutoff frequency of $5\pi/B$.

Tables I and II list SAD and AAD generated by the five methods in terms of Gaussian noise with different SNR levels and correlated noise, respectively. In order to compare performance conveniently, we label the best results in bold typeface at each designated SNR value, and the second best results are underlined. An identical operation is done for Tables III and IV in the latter experiments on real data sets. We can find that the performances of all methods are improved with the decrease in Gaussian noise corruption. Most importantly, the proposed method performs better and more stable than the others, indicating that it not only achieves more desired results, but also is more robust to noise contamination.

Furthermore, to intuitively illustrate the performance of SDNMF-TV, we take the case of SNR $= 20$ dB and $L = 3$ as an instance. Fig. 7 shows the USGS library spectral signatures and the estimated ones by SDNMF-TV on the synthetic data. From this figure, it is apparent that the estimated waveform and the ground truth coincide very well for each material. Meanwhile, the true abundance maps and the estimated ones by SDNMF-TV for seven endmembers are shown in Figs. 8 and 9, respectively. We can find that the distribution of each minerals is basically consistent with the true fractional abundance maps. Besides, in order to further illustrate the importance of the TV regularizer, we display the results achieved by SDNMF, as shown in Fig. 10. It can be seen that the noise points appeared in Fig. 9 are less than that in Fig. 10.

B. Real Data Experiments

1) Samson Data Set: The Samson data set is used as the first real hyperspectral data set [56], as shown in Fig. 11(a). It contains 952×952 pixels, and each pixel has 156 bands

TABLE III

COMPARISON OF THE SAD ON THE SAMSON DATA SET FOR DIFFERENT METHODS. WE LABEL THE BEST RESULTS FOR EACH ENDMEMBER IN BOLD TYPEFACE, AND THE SECOND-BEST RESULTS ARE UNDERLINED

Methods	SDNMF-TV	SDNMF	MLNMF	$L_{1/2}$-NMF	VCA-FCLS
Soil	0.0201	**0.0092**	0.0289	0.0270	0.0610
Tree	0.0297	**0.0283**	0.0490	0.0551	0.0495
Water	**0.0961**	0.1287	0.1290	0.1288	0.1299
Mean	**0.0486**	0.0554	0.0690	0.0703	0.0801

TABLE IV

COMPARISON OF THE SAD ON THE AVIRIS CUPRITE DATA SET FOR DIFFERENT METHODS. WE LABEL THE BEST RESULTS FOR EACH ENDMEMBER IN BOLD TYPEFACE, AND THE SECOND-BEST RESULTS ARE UNDERLINED

Methods	SDNMF-TV	SDNMF	MLNMF	$L_{1/2}$-NMF	VCA-FCLS
Alunite GDS82 Na82	0.0980	**0.0653**	0.1021	0.1131	0.1664
Andradite WS487	**0.0673**	0.1055	0.0686	0.1281	0.1259
Buddingtonite GDS85 D-206	0.1141	0.1543	**0.0993**	0.1606	0.1525
Chalcedony CU91-6A	0.1477	0.1431	0.1461	**0.1098**	0.1137
Kaolin/Smect H89-FR-5 30K	**0.0649**	0.0662	0.0676	0.0870	0.0878
Kaolin/Smect KLF508 85%K	**0.0726**	0.0772	0.1084	0.1476	0.1519
Kaolinite KGa-2	0.1075	**0.1072**	0.2203	0.1502	0.1654
Montmorillonite+Illi CM37	0.0800	**0.0485**	0.0519	0.0532	0.0502
Muscovite IL107	0.0989	0.1316	**0.0695**	0.1201	0.0876
Nontronite NG-1.a	0.1277	0.1213	**0.0809**	0.1308	0.1235
Pyrope WS474	0.1118	0.0951	0.1110	**0.0939**	0.1243
Sphene HS189.3B	0.0642	0.0692	0.0780	**0.0599**	0.0633
Mean	**0.0962**	0.0987	0.1003	0.1128	0.1177

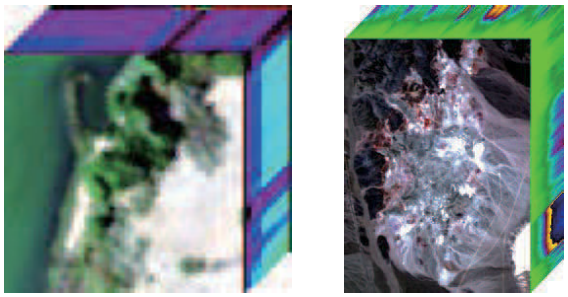


Fig. 11. (a) Subscene of Samson. (b) AVIRIS Cuprite image.

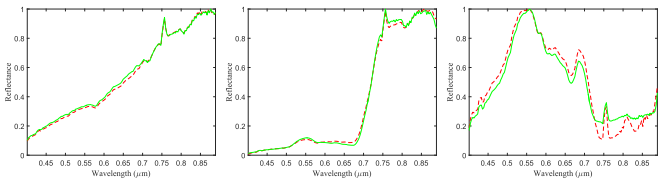


Fig. 12. Comparison of the USGS library spectra (green solid line) with the endmember signatures estimated by SDNMF-TV (red dashed line) on the Samson data set. (a) Soil. (b) Tree. (c) Water.

ranging from 0.401 to 0.889 μm. Due to the large size of the original image, a region of 95 × 95 pixels [with the pixel at spatial coordinates (252, 332) in the original image adopted as the upper leftmost pixel in the subscene] is considered to alleviate the computational burden. Based on the previous works [24], [30], there are three endmembers in the subscene: "#1 Soil", "#2 Tree", and "#3 Water".

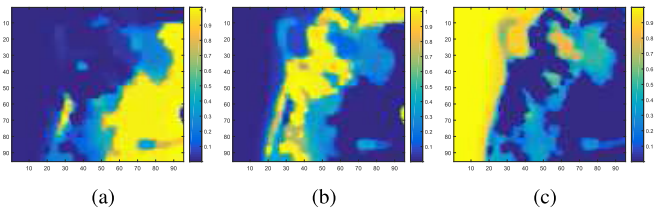


Fig. 13. Fractional abundance maps of three endmembers estimated by SDNMF-TV on the Samson data set. (a) Soil. (b) Tree. (c) Water.

The comparison of the reference spectra with the estimated signatures by SDNMF-TV on the Samson data set is shown in Fig. 12. It is apparent that the estimated spectra of the three materials are in accordance with the ground truth, revealing the effectiveness of the proposed method. Since the true abundances of the real data set are unavailable, Fig. 13 only displays the estimated abundance maps by SDNMF-TV.

Table III lists the SAD values yielded by our proposed method as well as three comparative methods for each material. From Table III, it can be easily found that the signature of water estimated by SDNMF-TV is better than that of others, and the mean value about SAD of SDNMF-TV is the smallest. This reveals that the proposed method outperforms other state-of-the-art algorithms.

2) AVIRIS Cuprite Data Set: The AVIRIS Cuprite data set acquired in 1995 is utilized as the second real HSI, as shown in Fig. 11(b). It contains 250 × 191 pixels, and each pixel has 224 bands ranging from 0.4 to 2.5 μm. The noisy channels (1 and 2 and 221–224) and water absorption channels (104–113 and 148–167) are removed, and 188 channels are remained. In addition, there are

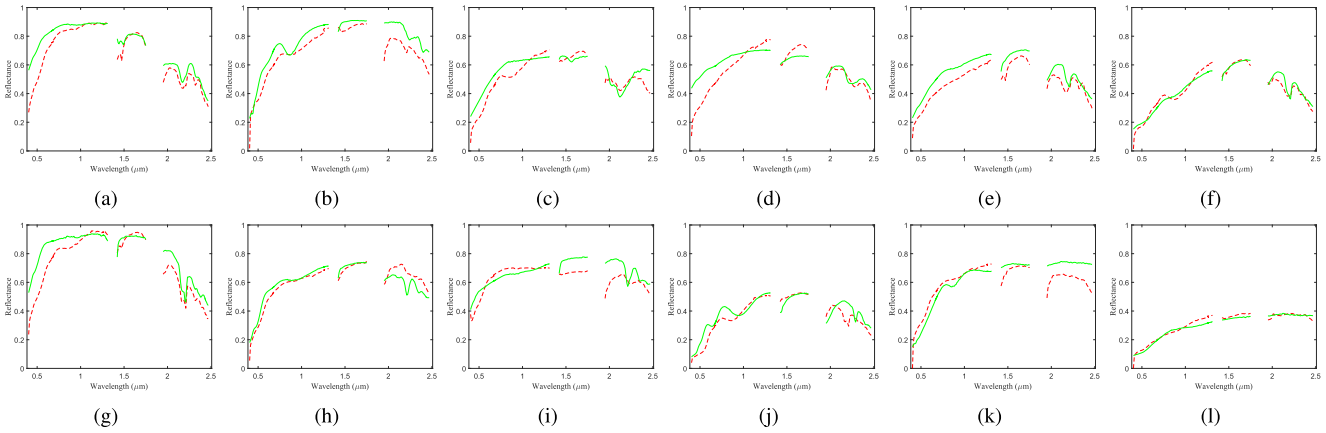


Fig. 14. Comparison of the USGS library spectra (green solid line) with the endmember signatures estimated by SDNMF-TV (red dashed line) on the AVIRIS Cuprite subscene. (a) Alunite GDS82 Na82. (b) Andradite WS487. (c) Buddingtonite GDS85 D-206. (d) Chalcedony CU91-6A. (e) Kaolin/Smect H89-FR-5 30K. (f) Kaolin/Smect KLF508 85%K. (g) Kaolinite KGa-2. (h) Montmorillonite + Illi CM37. (i) Muscovite IL107. (j) Nontronite NG-1.a. (k) Pyrope WS474. (l) Sphene HS189.3B.

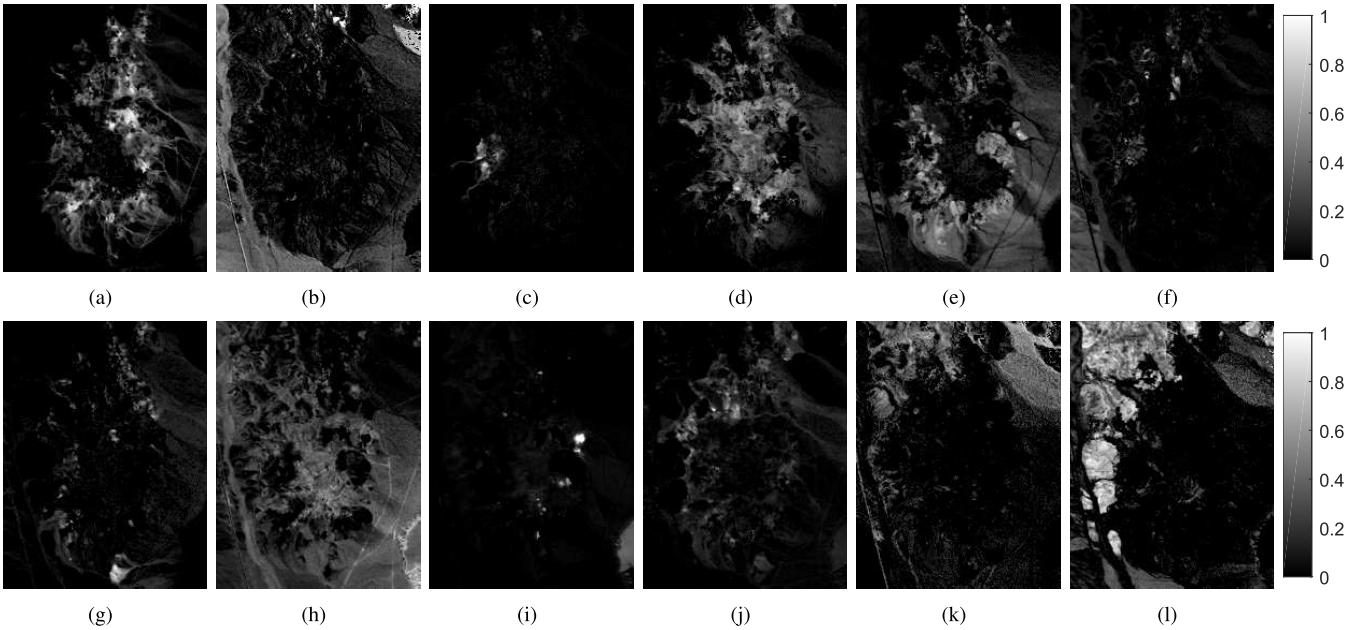


Fig. 15. Fractional abundance maps estimated by SDNMF-TV on the AVIRIS Cuprite subscene. (a) Alunite GDS82 Na82. (b) Andradite WS487. (c) Buddingtonite GDS85 D-206. (d) Chalcedony CU91-6A. (e) Kaolin/Smect H89-FR-5 30K. (f) Kaolin/Smect KLF508 85%K. (g) Kaolinite KGa-2. (h) Montmorillonite + Illi CM37. (i) Muscovite IL107. (j) Nontronite NG-1.a. (k) Pyrope WS474. (l) Sphene HS189.3B.

mainly 12 minerals: "#1 Alunite GDS82 Na82", "#2 Andradite WS487", "#3 Buddingtonite GDS85 D-206", "#4 Chalcedony CU91-6A", "#5 Kaolin/Smect H89-FR-5 30K", "#6 Kaolin/Smect KLF508 85%K", "#7 Kaolinite KGa-2", "#8 Montmorillonite + Illi CM37", "#9 Muscovite IL107", "#10 Nontronite NG-1.a", "#11 Pyrope WS474", and "#12 Sphene HS189.3B".

In Fig. 14, we compare the reference signatures selected from the library with the endmembers estimated by SDNMF-TV on the AVIRIS Cuprite data set. Similarly, our proposed method yields high spatial consistency of the dozen considered minerals, indicating that the method is effective. Besides, the abundance maps estimated by SDNMF-TV are shown in Fig. 15, where the brighter pixel explains higher abundance of the corresponding endmembers.

The corresponding SAD values for each mineral signature are given in Table IV. Compared with other considered methods, it can be observed that several minerals (e.g., Andradite WS487, Kaolin/Smect H89-FR-5 30K, and Kaolin/Smect KLF508 85%K) are estimated with the best performance for SDNMF-TV. Most importantly, the mean value about SAD of our proposed method is the smallest, which confirms the superiority of our proposed approach in this context.

VI. Conclusion

In this paper, a new SDNMF-TV has been proposed for linear spectral unmixing. First, we extend the NMF to a deep NMF model by adopting the concept of DL. The proposed model is able to learn hierarchical features with hidden information, consisting of *pretraining stage* and *fine-tuning stage*.

The former stage is performed to pretrain all factors layer by layer, and the latter stage fine-tunes all factors to reduce the total reconstruction error. On this basis, two constraints are enforced to exploit the spectral and spatial information contained in the original hyperspectral data. Specifically, we first introduce the $L_{1/2}$ regularizer to enhance the sparsity of the abundance matrix, and then, the TV model is further incorporated by considering piecewise smoothness of the abundances to build SDNMF-TV model. Besides, the newly proposed unmixing method is solved by multiplicative iterative algorithms. A detailed comparison with three competing unmixing approaches (i.e., MLNMF, $L_{1/2}$-NMF, and VCA-FCLS) demonstrates that our proposed method offers the best performance in the considered experiments. Future work will be focused on the extension of the proposed unmixing method to the hyperspectral data with the mixed noise (i.e., sparse noise and Gaussian noise).

REFERENCES

[1] 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.

[2] X. Bai, Z. Guo, Y. Wang, Z. Zhang, and J. Zhou, "Semisupervised hyperspectral band selection via spectral–spatial hypergraph model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2774–2783, Jun. 2015.

[3] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 Site," *J. Geophys. Res.*, vol. 91, no. B8, pp. 8098–8112, Jul. 1986.

[4] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, Sep. 2016.

[5] X. Bai, H. Zhang, and J. Zhou, "VHR object detection based on structural feature extraction and query expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6508–6520, Oct. 2014.

[6] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.

[7] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2017.

[8] 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.

[9] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.

[10] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.

[11] N. Dobigeon, J. Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[12] J. M. P. Nascimento and J. M. Bioucas-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.

[13] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. WHISPERS*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.

[14] Y. Xu, F. Fang, and G. Zhang, "Similarity-guided and ℓ_p-regularized sparse unmixing of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2311–2315, Nov. 2015.

[15] H. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4257–4266, Feb. 2016.

[16] R. Wang, H.-C. Li, W. Liao, X. Huang, and W. Philips, "Centralized collaborative sparse unmixing for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1949–1962, May 2017.

[17] R. Wang, H.-C. Li, A. Pizurica, J. Li, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using double reweighted sparse regression and total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1146–1150, Jul. 2017.

[18] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[19] J. D. Bayliss, J. A. Gualtieri, and R. F. Cromp, "Analyzing hyperspectral data with independent component analysis," in *Proc. SPIE 26th AIPR Workshop, Exploiting New Image Sources Sensors*, Washington, DC, USA, vol. 3240, Mar. 1998, pp. 133–143.

[20] W. Xia, X. Liu, B. Wang, and L. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2165–2179, Jun. 2011.

[21] N. Wang, B. Du, L. Zhang, and L. Zhang, "An abundance characteristic-based independent component analysis for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 416–428, Jan. 2015.

[22] W. Wang, Y. Qian, and Y. Y. Tang, "Hypergraph-regularized sparse NMF for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 681–694, Feb. 2016.

[23] L. Tong, J. Zhou, Y. Qian, X. Bai, and Y. Gao, "Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6531–6544, Nov. 2016.

[24] Y. E. Salehani and S. Gazor, "Smooth and sparse regularization for NMF hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3677–3692, Aug. 2017.

[25] W. Wang and Y. Qian, "Adaptive $L_{1/2}$ sparsity-constrained NMF with half-thresholding algorithm for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2618–2631, Jun. 2015.

[26] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.

[27] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[28] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained ℓ_p–ℓ_2 optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142–1146, Sep. 2013.

[29] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral unmixing with l_q regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6793–6806, Nov. 2014.

[30] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5412–5427, Dec. 2014.

[31] Z. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang, "$L_{1/2}$ regularization," *Sci. China Inf. Sci.*, vol. 53, no. 6, pp. 1159–1169, Jun. 2010.

[32] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2746–2758, May 2014.

[33] L. Tong, J. Zhou, X. Li, Y. Qian, and Y. Gao, "Region-based structure preserving nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1575–1588, Apr. 2017.

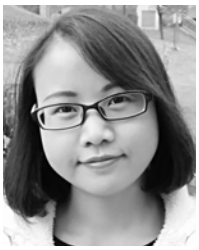
[34] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[35] S. Yang, X. Zhang, Y. Yao, S. Cheng, and L. Jiao, "Geometric nonnegative matrix factorization (GNMF) for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2696–2703, Jun. 2015.

[36] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

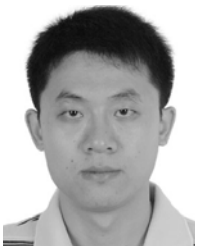
[37] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[38] G. A. Licciardi and F. D. Frate, "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4163–4172, Nov. 2011.
[39] J. Li, X. Li, B. Huang, and L. Zhao, "Hopfield neural network approach for supervised nonlinear spectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 1002–1006, Jul. 2016.
[40] A. Lemme, R. F. Reinhart, and J. J. Steil, "Online learning and generalization of parts-based image representations by non-negative sparse autoencoders," *Neural Netw.*, vol. 33, pp. 194–203, Sep. 2012.
[41] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *Proc. WHISPERS*, Tokyo, Japan, Jun. 2015, pp. 1–4.
[42] F. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Neural network hyperspectral unmixing with spectral information divergence objective," in *Proc. IGARSS*, Fort Worth, TX, USA, Jul. 2017, pp. 755–758.
[43] Y. C. Su, A. Marinoni, J. Li, A. Plaza, and P. Gamba, "Nonnegative sparse autoencoder for robust endmember extraction from remotely sensed hyperspectral images," in *Proc. IGARSS*, Fort Worth, TX, USA, Jul. 2017, pp. 205–208.
[44] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[45] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[46] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.
[47] W. Xie and Y. Li, "Hyperspectral imagery denoising by deep learning with trainable nonlinearity function," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1963–1967, Nov. 2017.
[48] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, Mar. 2017.
[49] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Nature Inf. Process. Syst.*, 2001, pp. 556–562.
[50] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
[51] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[52] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[53] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[54] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 38–42, Jan. 2015.
[55] W. Tang, Z. Shi, and Z. An, "Nonnegative matrix factorization for hyperspectral unmixing using prior knowledge of spectral signatures," *Opt. Eng.*, vol. 51, no. 8, p. 087001, Aug. 2012.
[56] *The Samson Data Set*. Accessed: Apr. 5, 2017. [Online]. Available: http://www.escience.cn/people/feiyunZHU/Dataset_GT.html



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.



Heng-Chao Li (S'06–M'08–SM'14) received the B.Sc. and M.Sc. degrees from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering.

From 2013 to 2014, he was a Visiting Scholar along with Prof. W. J. Emery at the University of Colorado, Boulder, CO, USA. He is currently a Professor with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University. His research interests include the statistical analysis of synthetic aperture radar images, remote sensing image processing, and signal processing in communications.

Dr. Li received several scholarships or awards, especially including the Special Grade of the Financial Support from the China Postdoctoral Science Foundation in 2009 and the New Century Excellent Talents in University from the Ministry of Education of China in 2011. He serves as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He has been a Reviewer for several international journals and conferences, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the *IET Radar, Sonar and Navigation*, the *IET Signal Processing*, the *IET Image Processing*, the *Pattern Recognition*, the *International Journal of Remote Sensing*, the *Remote Sensing*, and the *Canadian Journal of Remote Sensing*.

Jun Li (M'12–SM'16) was born in Loudi, Hunan, China, in 1982. She received the B.Sc. degree in geographical information systems from Hunan Normal University, Changsha, China, in 2004, the M.Sc. degree in remote sensing and photogrammetry from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical and computer engineering from the Instituto Superior Tecnico, Technical University of Lisbon, Lisbon, Portugal, in 2011.

From 2011 to 2012, she was a Post-Doctoral Researcher with the Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain. She is currently a Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China, where she founded her own research group on hyperspectral image analysis in 2013. Since 2013, she has been receiving several prestigious funding grants at the national and international levels. She has published a total of 69 journal citation report papers, 48 conference international conference papers, and one book chapter. Her research interests include remotely sensed hyperspectral image analysis, signal processing, supervised/semisupervised learning, and active learning.

Dr. Li's students have received important distinctions and awards at international conferences and symposia. She has served as a Guest Editor for a Special Issue in the prestigious PROCEEDINGS OF THE IEEE journal and a Special Issue in the prestigious *ISPRS Journal of Photogrammetry and Remote Sensing* journal. She has been serving as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2014. She has received a significant number of citations to her published works with several papers distinguished as Highly Cited Papers in Thomson Reuters Web of Science–Essential Science Indicators.

Qian Du (S'98–M'00–SM'05–F'18) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore, 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. She is also an Adjunct Professor with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China. 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 received the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was a Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013 and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She is the General Chair of the 4th IEEE Geoscience and Remote Sensing Society's Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Shanghai, in 2012. She has served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Antonio Plaza (M'05–SM'07–F'15) 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. He has authored over 600 publications, including 200 journal citation report papers (145 in IEEE journals), 23 book chapters, and 285 peer-reviewed conference proceeding papers. He has guest edited 10 special issues on hyperspectral remote sensing for different journals. He has reviewed over 500 papers for over 50 different journals. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was a member of the Editorial Board of the IEEE Geoscience and Remote Sensing Newsletter from 2011 to 2012 and the *IEEE Geoscience and Remote Sensing Magazine* in 2013. He was also a member of the steering Committee of the IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (JSTARS). He is a fellow of IEEE for contributions to hyperspectral data processing and parallel computing of earth observation data. He was a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) in 2010. He was a recipient of the Best Column Award from the *IEEE Signal Processing Magazine* in 2015, the 2013 Best Paper Award from the JSTARS journal, and the Most Highly Cited Paper Award (2005–2010) from the *Journal of Parallel and Distributed Computing*. He received the Best Paper Award from the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He served as an Associate Editor for the IEEE TGRS from 2007 to 2012. He is an Associate Editor of the IEEE ACCESS. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012 and the President for the Spanish Chapter of the IEEE GRSS from 2012 to 2016. He has served as the Editor-in-Chief for the IEEE TGRS from 2013 to 2017.

William J. Emery (M'90–SM'01–F'02) received the Ph.D. degree in physical oceanography from the University of Hawaii, Honolulu, HI, USA, in 1975.

After he was with Texas A&M University, College Station, TX, USA, he moved to The University of British Columbia, Vancouver, BC, Canada, in 1978, where he created a Satellite Oceanography Facility/Education/Research Program. In 1987, he was appointed as a Professor of aerospace engineering sciences with the University of Colorado, Boulder, CO, USA. He is an Adjunct Professor of informatics with Tor Vergata University, Rome, Italy. He has authored over 182 refereed publications, two textbooks, and over 150 conference papers.

Dr. Emery was an Elected Fellow of the American Meteorological Society in 2010, the American Astronautical Society in 2011, and the American Geophysical Union in 2012. He received the 2004 IEEE Geoscience and Remote Sensing Society (GRSS) Educational Award and the 2009 GRSS Outstanding Service Award. He was recently elected as the Chair of the IEEE Periodicals Committee which oversees the publication of 171 journals. He is the Vice President of Publications of the IEEE GRSS. He was the Founding Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2004 and the Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS for six years. He serves as the Editor-in-Chief of the *Journal of Atmospheric and Oceanic Technology*.