Deep Autoencoders with Multi-Task Learning for Bilinear Hyperspectral Unmixing

Yuanchao Su, *Member, IEEE,* Xiang Xu, *Member, IEEE,* Jun Li, *Senior Member, IEEE,* Hairong Qi, *Fellow, IEEE* Paolo Gamba, *Fellow, IEEE* Antonio Plaza, *Fellow, IEEE*

***Abstract*— Hyperspectral unmixing is an important problem for remotely sensed data interpretation. It amounts at estimating the spectral signatures of the pure spectral constituents in the scene (endmembers) and their corresponding sub-pixel fractional abundances. Although the unmixing problem is inherently nonlinear (due to multiple scattering), nonlinear unmixing of hyperspectral data has been a very challenging problem. This is because nonlinear models require detailed knowledge about the physical interactions between the sunlight scattered by multiple materials. In turn, bilinear mixture models (BMMs) can reach good accuracy with a relatively simple model for scattering. In this paper, we develop a new BMM and a corresponding unsupervised unmixing approach which consists of two main steps. In the first step, a deep autoencoder is used to linearly estimate the endmember signatures and their associated abundance fractions. The second step refines the initial (linear) estimates using a bilinear model, in which another deep autoencoder (with a low-rank assumption) is adopted to model second-order scattering interactions. It should be noted that in our developed BMM model, the two deep autoencoders are trained in a mutually interdependent manner under the multi-task learning framework, and the relative reconstruction error is used as the stopping criterion. The effectiveness of the proposed method is evaluated using both synthetic and real hyperspectral datasets. Our experimental results indicate that the proposed approach can reasonably estimate the nature of nonlinear interactions in real scenarios. Compared with other state-of-the-art unmixing algorithms, the proposed approach demonstrates very competitive performance.**

***Index Terms*—Hyperspectral nonlinear unmixing, Bilinear mixture, Autoencoder, Deep learning, Multi-task learning.**

This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences No. XDA19090104; National Natural Science Foundation of China under Grant No.61771496; Guangdong Provincial Natural Science Foundation under Grant No.2016A030313254; National Key Research and Development Program of China under Grant No.2017YFB0502900; Social Welfare Research Project of Zhongshan City under Grant NO.2018B1015 and No.2019B2026; FEDER/Junta de Extremadura GR18060; and the European Union's Horizon 2020 research and innovation programme under grant agreement 734541 EOXPOSURE. (Corresponding author: Xiang Xu.)

Y. Su is with the Department of Remote Sensing, College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054, China (e-mail: suych3@xust.edu.cn).

X. Xu is with the Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528402, China (e-mail: xuxiang@zsc.edu.cn).

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou, 510275, China (e-mail:lijun48@mail.sysu.edu.cn).

H. Qi is with the Advanced Imaging and Collaborative Information Processing Group, Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA (e-mail: hqi@utk.edu).

P. Gamba is with the Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia, Pavia, I-27100, Italy (e-mail: paolo.gamba@unipv.it).

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, E-10071, Spain (e-mail: aplaza@unex.es).

I. Introduction

Hyperspectral remote sensing, which combines traditional imaging and physical (spectral) analysis techniques, has been widely used for Earth observation purposes [1]. Hyperspectral sensors can acquire fine spectral information and rich spatial details of the target scenes, thus achieving strong discrimination ability for different land-cover classes and promoting the development of quantitative remote sensing [2], [3]. Nowadays, hyperspectral imagery have been readily used in many geoscience fields, including environmental monitoring, precision agriculture and land-use surveying [4], [5]. However, due to the relatively low spatial resolution of hyperspectral images, plenty of pixels are mixed by several different substances, leading to inaccuracies in the understanding and quantification of the considered scenes [6]. To address this issue, spectral unmixing (SU) has been adopted to decompose each pixel spectrum into a collection of pure constituent spectra (called endmembers), and their corresponding abundance fractions [7]. Most SU approaches are based on linear or nonlinear spectral mixture models [8]–[10].

The linear mixture model (LMM) assumes that the endmembers interact linearly with the incident radiation at a sub-pixel level [11]. Many approaches have been developed under the LMM assumption. These algorithms generally perform unmixing without considering the scattering interactions between materials. Among many examples, we can mention N-FINDR [12], vertex component analysis (VCA) [13], piecewise convex multiple-model endmember detection [14], minimum volume simplex analysis [15], joint Bayesian algorithm [16], incremental proximal sparse and low-rank unmixing [17], subspace unmixing with low-rank attribute [18], robust collaborative non-negative matrix factorization [19], $L_{1/2}$-NMF [20], and sparsity-constrained deep NMF [21].

In recent years, autoencoders have also been widely used for developing robust SU approaches based on the LMM. Autoencoders, as a learning model based on artificial neural networks, provide a combination of an encoder and a decoder to carry out unsupervised learning [22]. The encoder transforms the input data into a code, and the decoder reconstructs the input data from the corresponding code [23]. By means of a shared weight strategy between adjacent layers (as well as contrastive divergence), autoencoders can be trained quite efficiently [24]. For instance, the marginalized denoising autoencoder and a

nonnegative sparse autoencoder were combined in the form of a cascade method to conduct SU, exhibiting robustness and intrinsic self-adaptation capabilities [25]. Also, several other autoencoder-based SU methods have been developed, including the stacked nonnegative sparse autoencoder [26], untied denoising autoencoder [27], sparse autoencoder network [28], deep autoencoder network [29], and a neural network autoencoder [30].

However, all the aforementioned LMM models are based on a rough assumption that may not hold in practice [31]. Nonlinear mixture models (NLMMs), instead, aim at characterizing the interactions between the sunlight scattered by multiple materials, being at an intimate or multilayered level [32], [33].

- For intimate mixtures, the interactions consist of photons emitted from molecules of one material absorbed by molecules of another material, which may in turn emit more photons [34]–[37]. Generally speaking, the models that assume intimate mixtures are more faithful to the imaging mechanism of the instrument. However, it is worth noting that these models are strongly sensitive to some inherent parameters, which are often determined by the knowledge of the geometric positioning of the sensor. Such dependency upon external parameters brings great difficulties when implementing the inversion [32].
- For multilayered mixtures, the corresponding models hold that multiple interactions among the scatterers are at different layers, and lead to an infinite sequence of powers of products of reflectance, which are formalized to a combination of the LMM and additional high-order terms [31], [38]. In this case, high-order scattering (derived from the high-order interactions of photons between different materials), is negligibly weak.

NLMM approaches generally estimate the impact of the scattering interactions, such as the generalized bilinear model (GBM) unmixing [39], semi-nonnegative matrix factorization (Semi-NMF) [40], or robust nonnegative matrix factorization (rNMF) [41]. The GBM and Semi-NMF are supervised methods which need to know endmembers in advance, while rNMF is a fully unsupervised algorithm. However, the nonlinear interactions on abundances still cannot be obtained by rNMF. Moreover, many kernel-based nonlinear unmixing methods have been developed by using kernel functions and physically-inspired models [42], [43], opening up new avenues for nonlinear unmixing [31].

Unfortunately, most NLMM approaches are too complex and difficult to exploit in real scenarios, since they require detailed knowledge about the physical interactions between the sunlight scattered by multiple materials. For simplicity, bilinear mixture models (BMMs) have been proposed and used more commonly in practice; these models consider the second-order scattering of photons between two materials [44]. Despite the fact that they neglect high-order scattering, BMMs can provide very good approximations to the sunlight scattering mechanism in real scenarios [32]. Under the BMM, the formula of SU can be regarded as a combination of linear and bilinear terms, in which the linear components explain endmembers and abundances, and bilinear components reflect the second-order scattering interactions [39].

The development of methods able to model both linear and nonlinear interactions is quite challenging, as they involve different tasks in the optimization procedure [45]. In general, any optimization task is based on a training procedure in which the model is iteratively trained (with a particular metric) until it converges to a stable solution [46]. However, if the model only focuses on one single task, it generally ignores some information which may be helpful for the learning of the other related tasks [47]. To solve the above problem, multi-task learning (MTL) was developed to improve the optimization performance by utilizing shared information among different tasks [48], [49]. As a common optimization strategy, MTL has been successfully applied in many fields, such as speech recognition, computer vision, and natural language processing [48], [50], and also have wide application prospect in the hyperspectral unmixing. Recently, a linear unmixing approach, by using MTL and spectral-spatial information, has been introduced in [51], demonstrating a good potential for hyperspectral unmixing.

In this paper, we develop an unsupervised unmixing approach based on the hierarchical BMM, which can estimate the endmember signatures, the abundance fractions, the interaction between abundance fractions, and the interaction outliers caused by bilinear components. Here, the interaction outliers physically represent the scattering interactions between materials in a hyperspectral image, simultaneously. To do so, our method operates within a MTL framework to combine two different tasks. Task 1 estimates endmembers and abundances via a deep autoencoder, while task 2 adopts another deep autoencoder to update the bilinear components, simultaneously. The encoders are able to decompose hyperspectral data and the decoders can reconstruct the data. It should be noted that, the two deep autoencoders are trained in a mutually interdependent manner, and the stopping criterion is based on the relative reconstruction error (RE). Besides, we also integrate some aspects of the deep semi-NMF [52] into the aforementioned autoencoders, in order to learn some hidden representations from the data and further improve the performance of optimization.

Compared with other nonlinear unmixing algorithms, the main advantages of our newly proposed method can be summarized as follows:

1) We provide a fully unsupervised unmixing approach that does not require any prior knowledge.
2) Its unmixing performance is enhanced by adopting a hierarchical framework, able to learn multiple hidden representations.
3) By taking advantage of the MTL framework, our unmixing (based on the BMM) is split into two tasks which can be executed in parallel.
4) Our method can estimate endmembers and abundances, as well as the interaction abundances and interaction outliers, simultaneously.

The remainder of this paper is organized as follows. In section II, we review some related works. Section III describes the proposed method in detail. In section IV, synthetic and real hyperspectral images are used for evaluation purposes,

allowing us to conduct a quantitative comparison with other state-of-the-art unmixing algorithms. Section V concludes the paper with some remarks and hints at plausible future research.

II. Related Works

Our proposed approach employs two deep autoencoders to execute blind source separation on the observed data, which will bring multiple hidden representations associated with the BMM. The linear and bilinear components in the BMM correspond to the two hierarchies, respectively.

A. BMM

Let $\mathbf{Y} \equiv [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{R}^{d\times n}$ be a hyperspectral image with n spectral pixels and d spectral bands. In this work, the constraints of the BMM are derived from the Fan model [53]. For a hyperspectral image, the matrix form of the BMM is denoted as

$$\begin{aligned} &\mathbf{Y} = \mathbf{EA} + \mathbf{DB} + \mathbf{N}, \\ &\text{s.t. } \mathbf{A} \geq \mathbf{0}, \mathbf{B} \geq \mathbf{0}, \mathbf{1}_c^\top \mathbf{A} = \mathbf{1}_n^\top, \end{aligned} \tag{1}$$

where $\mathbf{E} \equiv [\mathbf{e}_1, \ldots, \mathbf{e}_c] \in \mathbb{R}^{d\times c}$ represents the mixing matrix with c endmembers, $\mathbf{A} \equiv [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbb{R}^{c\times n}$ denotes abundance matrix. For this product term, the two constraints $\mathbf{A} \geq 0$ and $\mathbf{1}_c^\top \mathbf{A} = \mathbf{1}_n^\top$ are commonly known as the abundance nonnegativity constraint (ANC) and abundance sum-to-one constraint (ASC) with $\mathbf{1}_c = [1, 1, \ldots, 1]^\top \in \mathbb{R}^c$, which stem from a physical interpretation of the abundance fractions [54]. The matrix $\mathbf{D} \equiv [\mathbf{d}_1, \ldots, \mathbf{d}_l] \in \mathbb{R}^{d\times l}$ contains l virtual endmembers, and $\mathbf{B} \equiv [\mathbf{b}_1, \ldots, \mathbf{b}_l] \in \mathbb{R}^{l\times n}$ denotes an interaction abundance matrix, with $l = c(c-1)/2$. The matrix $\mathbf{N} \in \mathbb{R}^{d\times n}$ denotes an error matrix that may affect the imaging process (e.g. noise). Note that the notation $[\cdot]^\top$ stands for the transposition of a matrix or vector. Following [40], we assume that a virtual endmember $\mathbf{d}_{(i,j)}$ is initialized by the corresponding endmembers. With these assumptions in mind, we have

$$\mathbf{d}_{(i,j)} = \mathbf{e}_i \odot \mathbf{e}_j, \tag{2}$$

where $\odot$ is the Hadamard product, $\mathbf{e}_i$ is the i-th endmember signature in $\mathbf{E}$, $\forall\, i \in \{1, \ldots, c-1\}$, $j \in \{i+1, \ldots, c\}$. Let $b_{(i,j),q}$ be an interaction abundance, initialized by

$$b_{(i,j),q} = a_{i,q} a_{j,q}, \tag{3}$$

where $b_{(i,j),q}$ is an element of $\mathbf{b}_q$ at the q-th column of $\mathbf{B}$, and $a_{i,q}$ and $a_{j,q}$ are the i-th element and the j-th element of the q-th column of $\mathbf{A}$, respectively.

B. Hierarchical BMM

In (1), let $\mathbf{Y}_L \in \mathbb{R}^{d\times n}$ be the linear term with

$$\mathbf{Y}_L = \mathbf{EA},$$

and let $\mathbf{Y}_N \in \mathbb{R}^{d\times n}$ be the bilinear term with

$$\mathbf{Y}_N = \mathbf{DB},$$

. Following [52], [55], [56], we have

$$\begin{aligned} \mathbf{Y}_L &= \mathbf{E}_1 \mathbf{E}_2 \cdots \mathbf{E}_m \mathbf{A}_m \\ \mathbf{Y}_N &= \mathbf{D}_1 \mathbf{D}_2 \cdots \mathbf{D}_m \mathbf{B}_m, \end{aligned} \tag{4}$$

respectively, where m is the number of hidden layers of the encoder or decoder. For the autoencoders used, the structure of the encoder is the same as that of the decoder. In the hierarchy, $\mathbf{E}_1, \cdots, \mathbf{E}_m$ denote weight matrices between adjacent layers, and $\mathbf{A}_1, \cdots, \mathbf{A}_m$ represent hidden-layers. Similarly, $\mathbf{D}_1, \cdots, \mathbf{D}_m$ are weight matrices, and $\mathbf{B}_1, \cdots, \mathbf{B}_m$ represent hidden-layers. Note that the sizes of these matrices may be not same, $\mathbf{E}_1 \in \mathbb{R}^{d\times c}$, $\mathbf{D}_1 \in \mathbb{R}^{d\times l}$, $\mathbf{E}_k \in \mathbb{R}^{c\times c}$, $\mathbf{D}_k \in \mathbb{R}^{l\times l}$, with $k = 2, \ldots, m$. Following [52], the hierarchies for the multiple hidden representations are given as

$$\begin{aligned} \mathbf{A}_{m-1} &\approx \mathbf{E}_m \mathbf{A}_m \\ &\vdots \\ \mathbf{A}_2 &\approx \mathbf{E}_3 \cdots \mathbf{E}_m \mathbf{A}_m \\ \mathbf{A}_1 &\approx \mathbf{E}_2 \cdots \mathbf{E}_m \mathbf{A}_m, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \mathbf{B}_{m-1} &\approx \mathbf{D}_m \mathbf{B}_m \\ &\vdots \\ \mathbf{B}_2 &\approx \mathbf{D}_3 \cdots \mathbf{D}_m \mathbf{B}_m \\ \mathbf{B}_1 &\approx \mathbf{D}_2 \cdots \mathbf{D}_m \mathbf{B}_m. \end{aligned} \tag{6}$$

For the linear term, $\mathbf{Y}_L$ is first decomposed into $\mathbf{E}_1$ and $\mathbf{A}_1$, $\mathbf{Y}_L \approx \mathbf{E}_1 \mathbf{A}_1$, and the hidden layers are then decomposed one-by-one, $\mathbf{A}_q \approx \mathbf{E}_{q+1} \mathbf{A}_{q+1}$, $\mathbf{A}_q \in \mathbb{R}^{c\times n}$, $q = 2, \ldots, m-1$. In the same way, the bilinear term is expressed as $\mathbf{Y}_N \approx \mathbf{D}_1 \mathbf{B}_1$, $\mathbf{B}_q \approx \mathbf{D}_{q+1} \mathbf{B}_{q+1}$, $\mathbf{B}_q \in \mathbb{R}^{l\times n}$. For the linear components, $\mathbf{E}$ and $\mathbf{A}$ are given by the following mathematical formulations:

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_1 \cdots \mathbf{E}_{m-1} \mathbf{E}_m \\ \mathbf{A} &= \mathbf{A}_m. \end{aligned} \tag{7}$$

In the same way as the bilinear components, $\mathbf{D}$ and $\mathbf{B}$ are given as follows:

$$\begin{aligned} \mathbf{D} &= \mathbf{D}_1 \cdots \mathbf{D}_{m-1} \mathbf{D}_m \\ \mathbf{B} &= \mathbf{B}_m. \end{aligned} \tag{8}$$

III. Proposed Method

In the proposed method, the autoencoders are trained together by minimizing the discrepancy between the original data and the reconstructed ones, and the objective optimization function is defined as follows:

$$(\mathbf{E}, \mathbf{A}, \mathbf{D}, \mathbf{B}) = \arg \min_{\mathbf{E},\mathbf{A},\mathbf{D},\mathbf{B}} \|\mathbf{Y} - \mathbf{EA} - \mathbf{DB}\|_F^2, \tag{9}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. The training of our method contains two stages: pre-training and fine-tuning. The two autoencoders first pre-train all factors layer-by-layer, respectively, and then all factors are fine-tuned. In the training, each node of a hidden layer in task 2 is activated by the Hadamard product between any two nodes of the corresponding hidden layer in task 1. Let $\widehat{\mathbf{Y}} = \mathbf{EA} + \mathbf{DB}$ be the reconstruction of the image $\mathbf{Y}$, $\widehat{\mathbf{Y}} \equiv [\widehat{\mathbf{y}}_1, \ldots, \widehat{\mathbf{y}}_n] \in \mathbb{R}^{d\times n}$.

The fine-tuning would stop when the reconstruction error (RE) reaches convergence, and the RE is denoted as

$$\mathrm{RE}\left(\{\mathbf{y}_g\}_{g=1}^n, \{\widehat{\mathbf{y}}_g\}_{g=1}^n\right) = \frac{1}{n}\sum_{g=1}^{n}\sqrt{\|\mathbf{y}_g - \widehat{\mathbf{y}}_g\|_2^2}. \quad (10)$$

The problem in (9) is nonconvex and therefore a global optimum is difficult to obtain in practice. Nevertheless, the proposed method can converge to a local optimum with reasonable computational cost. In the MTL framework, the two autoencoders correspond to two tasks, and each autoencoder has $2m$-1 hidden layers, because the encoder and decoder have the same structure, as illustrated in Fig. 1. In task 1, the endmembers are obtained from weight matrices of the decoder, and the abundance matrix corresponds to $\mathbf{A}_m$ in the middle hidden-layer. While in task 2, interaction outliers are estimated from the last layer of the decoder, and the interaction abundances correspond to $\mathbf{B}_m$ in the middle hidden-layer. For the autoencoder of the corresponding linear term, the number of nodes in each hidden layer is set to the number of endmembers. For the autoencoder of the corresponding bilinear term, the number of nodes in each hidden layer amounts to the number of virtual endmembers. The number of input and output nodes of the two autoencoders are all equal to the number of pixels in the scene.

A. Task 1 for Estimating Linear Components

Task 1 is used for estimating the endmember signatures and the abundance fractions, and the objective function of the sub-problem is written as

$$\mathcal{C}_{Task1} = \frac{1}{2}\left(\|\mathbf{Y} - \mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_m\mathbf{A}_m - \mathbf{DB}\|_F^2 + \mu\Re(\mathbf{A}_k)\right), \quad (11)$$

where μ is the regularization coefficient of Task 1. For this task, $\mathbf{D}$ and $\mathbf{B}$ are regarded as the constant matrices. With respect to $\mathbf{A}_k$, each of its rows corresponds to a nodes of the k-th hidden layer. Note that, although $\mathbf{A}_k$ is related to abundances, it is not directly equivalent to abundance fractions. Following [52], [57], $\Re(\mathbf{A}_k) = \mathrm{Tr}\left(\mathbf{A}_k\mathbf{L}\mathbf{A}_k^\top\right)$ controls the smoothness of $\mathbf{A}_k$, where $\mathrm{Tr}(\cdot)$ denotes the trace norm. Here, $\mathbf{L} = \mathbf{X} - \mathbf{W}$ is the graph Laplacian matrix obtained from the radial basis kernel function. $\mathbf{W}$ is called graph connection matrix, in which each element w_{ph} represents the graph connection between p-th pixel and h-th pixel, and satisfies $w_{ph} = w_{hp}$. $\mathbf{X}$ is a diagonal matrix, in which each diagonal element is the column sum of $\mathbf{W}$. In this autoencoder, all hidden-layers share $\mathbf{L}$, and the nodes associated with the abundances. Following [58], the graph connection is denoted as

$$w_{ph} = \begin{cases} \exp\left(-\frac{\|\mathbf{y}_p - \mathbf{y}_h\|^2}{2\sigma^2}\right), & \text{if } p \neq h \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where $\mathbf{y}_p$ and $\mathbf{y}_h$ are vectors of $\mathbf{Y}$, $\forall\, p \in \{1,\ldots,n-1\}$, $h \in \{p+1,\ldots,n\}$. The parameter σ can be estimated from a conjugate inverse Gamma distribution [59]. The generation of $\mathbf{W}$ is defined as follows:

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix}. \quad (13)$$

Summing up w_{ph} in each row of $\mathbf{W}$, the diagonal matrix $\mathbf{X}$ is obtained as

$$\mathbf{X} = \begin{bmatrix} w_{12}+\cdots w_{1n} & 0 & \cdots & 0 \\ 0 & w_{21}+\cdots w_{2n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{n1}+\cdots w_{n(n-1)} \end{bmatrix}. \quad (14)$$

In order to meet the ASC and accelerate convergence, we use the multiplicative update rule (MUR) to obtain $\mathbf{E}_1$ and $\mathbf{A}_1$, where the adopted MUR is formula from [60].

$$\begin{aligned} \mathbf{E}_1 &\leftarrow \mathbf{E}_1 \odot \left(\mathbf{Y}\mathbf{A}_1^\top\right) \cdot / \left(\mathbf{E}_1\mathbf{A}_1\mathbf{A}_1^\top\right) \\ \mathbf{A}_1 &\leftarrow \mathbf{A}_1 \odot \left(\widetilde{\mathbf{A}}_1^\top\widetilde{\mathbf{Y}}\right) \cdot / \left(\widetilde{\mathbf{E}}^\top\widetilde{\mathbf{E}}\mathbf{A}_1\right), \end{aligned} \quad (15)$$

where $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{A}}_1$ are the augmented matrices for satisfying the ASC constraint, and $\cdot/$ represents an element-wise division. As a result, we have

$$\widetilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \delta\mathbf{1} \end{bmatrix}, \quad \widetilde{\mathbf{A}}_1 = \begin{bmatrix} \mathbf{A}_1 \\ \delta\mathbf{1} \end{bmatrix}, \quad (16)$$

where δ is able to control the impact of the ASC constraint. For the deep autoencoder, δ can affect $\mathbf{A}_1$, and then it can also affect $\mathbf{A}_k$, because $\mathbf{A}_1$ is associated with $\mathbf{A}_k$. After evaluating the performance of our algorithm with different values of δ, we empirically set $\delta = 23$ in this work.

With respect to other weights and hidden layers, the gradient descent (GD) is employed to implement the optimization. The first-order partial derivatives of $\mathcal{C}_{Task1}$ with respect to $\mathbf{A}_k$ are defined as

$$\nabla_{\mathbf{A}_k}\mathcal{C}_{Task\ 1} = \mathbf{E}_k^\top\boldsymbol{\Psi}^\top(\boldsymbol{\Psi}\mathbf{E}_k\mathbf{A}_k - \mathbf{Y}) + \mu\mathbf{A}_k\mathbf{L}, \quad (17)$$

where $\boldsymbol{\Psi}$ is defined as $\boldsymbol{\Psi} = \mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_{k-1}$. Similarly, we have the partial derivative of $\mathcal{C}_{Task1}$ for $\mathbf{E}_k$.

$$\nabla_{\mathbf{E}_k}\mathcal{C}_{Task\ 1} = \boldsymbol{\Psi}^\top(\boldsymbol{\Psi}\mathbf{E}_k\widetilde{\mathbf{A}}_k - \mathbf{Y})\widetilde{\mathbf{A}}_k^\top, \quad (18)$$

where $\widetilde{\mathbf{A}}_k$ is a reconstruction, $\widetilde{\mathbf{A}}_k = \mathbf{E}_{k+1}\ldots\mathbf{E}_m\mathbf{A}_m$. The derivations of (17) and (18) are shown in an Appendix.

Afterwards, $\mathbf{E}_k$ and $\mathbf{A}_k$ are updated via the GD.

$$\begin{aligned} \mathbf{A}_k &\leftarrow \mathbf{A}_k - \eta_A\nabla_{\mathbf{A}_k}\mathcal{C}_{Task1} \\ \mathbf{E}_k &\leftarrow \mathbf{E}_k - \eta_E\nabla_{\mathbf{E}_k}\mathcal{C}_{Task1}, \end{aligned} \quad (19)$$

where the corresponding learning rates η_E and η_A are estimated by Armijo rule [61]. Finally, we obtain the abundances $\mathbf{A}$ and endmembers $\mathbf{E}$ as follows:

$$\begin{aligned} \mathbf{A} &= \mathbf{A}_m \\ \mathbf{E} &= \mathbf{E}_1\cdots\mathbf{E}_{m-1}\mathbf{E}_m. \end{aligned} \quad (20)$$

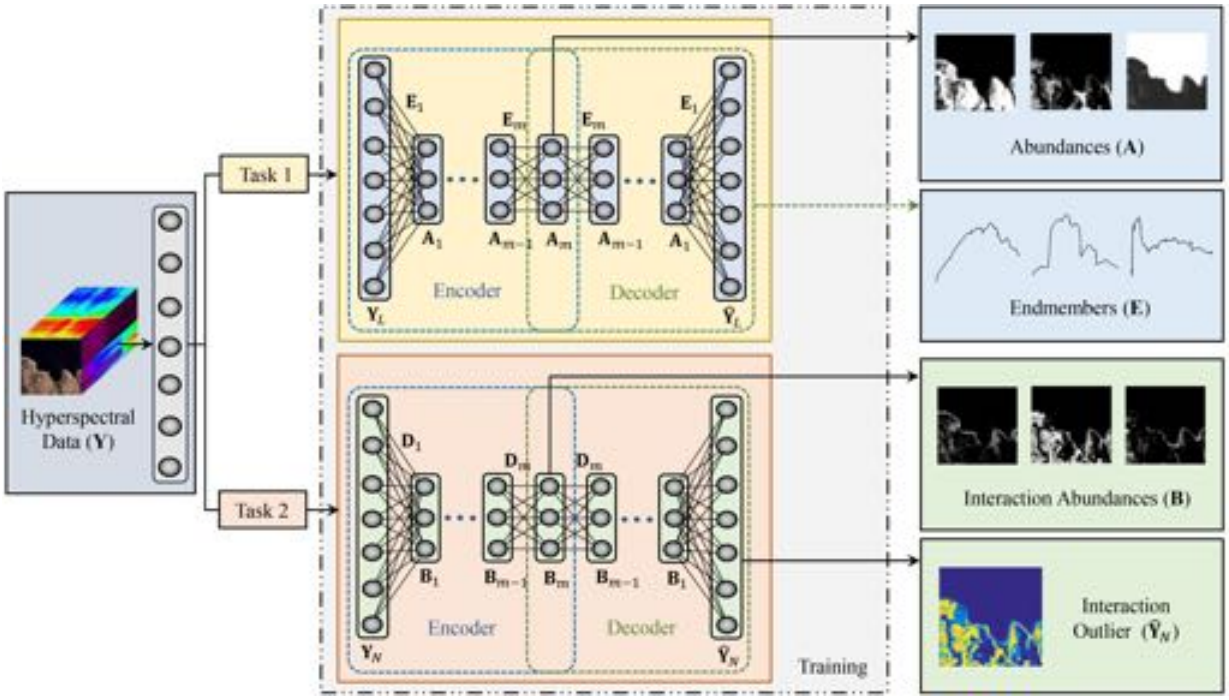


Fig. 1. Flowchart of the proposed method. Task 1 is used for updating the linear components, estimating the endmember signatures $\mathbf{E}$ and abundance fractions $\mathbf{A}$. Task 2 responds to the bilinear components, obtaining the interaction abundances $\mathbf{B}$ and the interaction outliers $\widehat{\mathbf{Y}}_N$.

B. Task 2 for Estimating Bilinear Components

Task 2 estimates the bilinear components, i.e., the interactions between abundance fractions and outlier. In hyperspectral images, the scattering interactions are mainly located in the boundaries and mixing regions between materials [41]. Therefore, compared with the material distributions, the scattering interactions usually have low-rank property. The objective function of the sub-problem is defined as

$$\mathcal{C}_{Task\ 2} = \frac{1}{2}\left(\|\mathbf{Y} - \mathbf{E}\mathbf{A} - \mathbf{D}_1\mathbf{D}_2\cdots\mathbf{D}_m\mathbf{B}_m\|_F^2 + \upsilon\|\mathbf{B}_m\|_*\right), \tag{21}$$

where $\|\mathbf{B}_m\|_* = \mathrm{Tr}(\sqrt{\mathbf{B}_m{}^\top\mathbf{B}_m})$ is used to impose low-rank, $\|\cdot\|_*$ denotes the nuclear norm, υ is the regularization coefficient. Following the update rule of the deep semi-NMF method in [52], we adopt it to update $\mathbf{B}_r \in \mathbb{R}^{l\times n}$.

$$\mathbf{B}_r \leftarrow \mathbf{B}_r \odot \sqrt{\frac{\left[\mathbf{\Phi}^\top(\mathbf{Y}-\mathbf{E}\mathbf{A})\right]^+ + \left[\mathbf{\Phi}^\top\mathbf{\Phi}\right]^-\mathbf{B}_r}{\left[\mathbf{\Phi}^\top(\mathbf{Y}-\mathbf{E}\mathbf{A})\right]^- + \left[\mathbf{\Phi}^\top\mathbf{\Phi}\right]^+\mathbf{B}_r}}, \tag{22}$$

where $\mathbf{\Phi} = \mathbf{D}_1\mathbf{D}_2\cdots\mathbf{D}_{r-1}$, $r = 1, ..., m-1$. Following [62], $[\cdot]^+$ and $[\cdot]^-$ are the positive and negative parts of the matrix, respectively. The update rule of $\mathbf{D}_r \in \mathbb{R}^{l\times l}$ is denoted as

$$\mathbf{D}_r \leftarrow \mathbf{D}_r \odot \left(\mathbf{\Phi}^\top(\mathbf{Y}-\mathbf{E}\mathbf{A})\mathbf{B}_r^\top\right) ./ \left(\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}_r\mathbf{B}_r\mathbf{B}_r^\top\right). \tag{23}$$

In task 2, the m-th hidden layer needs a constraint to ensure a low-rank and non-negative attribute on interaction abundances, thus we first set $\mathbf{D}_m = \mathbf{D}_{m-1}$ and $\mathbf{B}_m = \mathbf{B}_{m-1}$ to provide the initial values for $\mathbf{D}_m$ and $\mathbf{B}_m$. According to the constraint in [63], we adopt the following rule for $\mathbf{B}_m$.

$$\mathbf{B}_m \leftarrow \mathrm{sgn}(\widetilde{\mathbf{B}}_m) \odot \max\left(\widetilde{\mathbf{B}}_m - \mathbf{\Lambda}_{\widetilde{B}}, 0\right), \tag{24}$$

where $\mathrm{sgn}(\cdot)$ is the sign function, and the threshold is given as $\mathbf{\Lambda}_{\widetilde{B}} = \lambda./(|\widetilde{\mathbf{B}}_m| + \lambda)$ with $\lambda = 0.001$. The hyper-parameter λ avoids that the denominator is a zero matrix.

$$\widetilde{\mathbf{B}}_m = \mathbf{U}\ \mathrm{diag}(\widetilde{\mathbf{S}})\ \mathbf{V}^T, \tag{25}$$

where $\mathbf{U}, \mathbf{S}, \mathbf{V}$ are obtained by a singular value decomposition (SVD) of $\mathbf{B}_m$, i.e., $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \mathrm{SVD}(\mathbf{B}_m)$ and $\widetilde{\mathbf{S}} = \max(|\mathrm{diag}(\mathbf{S})| - \mathbf{\Lambda}_S, 0)$, with $\mathbf{\Lambda}_S = \lambda./(|\mathrm{diag}(\mathbf{S})| + \lambda)$. Let $\mathbf{D} = \mathbf{D}_1\cdots\mathbf{D}_{m-1}\mathbf{D}_m$. Then, we obtain the interaction abundances and interaction outliers as follows:

$$\begin{aligned}\mathbf{B} &= \mathbf{B}_m\\ \widehat{\mathbf{Y}}_N &= \mathbf{D}\mathbf{B}.\end{aligned} \tag{26}$$

C. Implementation Details

In this subsection, we provide the implementation details for our algorithm. First and foremost, we emphasize that we assume the number of endmembers to be known in advance. In addition, the weights and hidden-layers are pre-trained using the strategy in [21]. In task 1, $\mathbf{E}_1$ is obtained by running a pure pixel-based unmixing algorithm (e.g. VCA). Meanwhile, $\mathbf{A}_1$ is obtained by using the fully constrained least square (FCLS) method in [54]. Afterwards, $\mathbf{E}_k$ and $\mathbf{A}_k$ are initialized by using the same strategy. With respect to task 2, the weights and hidden-layers are initialized by using the corresponding factors resulting from task 1. More specifically, the initialization of

$\mathbf{d}_{(i,j),1}$ is calculated by $\mathbf{d}_{(i,j),1} = \mathbf{e}_{i,1} \odot \mathbf{e}_{j,1}$, where $\mathbf{e}_{i,1}$ and $\mathbf{e}_{j,1}$ are two column vectors in $\mathbf{E}_1$, $\forall\ i \in \{1, \ldots, c-1\}$, $j \in \{i+1, \ldots, c\}$. Then, $\mathbf{D}_1$ is composed by $\mathbf{d}_{(i,j),1}$. Here, we assume that $b_{(i,j),q,1}$ is an element of the q-th column of $\mathbf{B}_1$, $b_{(i,j),q,1} = a_{i,q,1}a_{j,q,1}$, where $a_{i,q,1}$ and $a_{j,q,1}$ are the i-th element and the j-th element of the q-th column of $\mathbf{A}_1$, respectively. Afterwards, $\mathbf{B}_1$ is composed by $b_{(i,j),q,1}$, The weights and nodes of other layers are then initialized using this same strategy. Finally, the fine-tuning of each layer is performed by minimizing the objective function in (9), until the RE in (10) achieves convergence. A pseudo-code of the proposed method is illustrated in Algorithm 1. In lines 1-2, the pre-training is implemented via VCA and FCLS. In lines 3-4, the weights and nodes of hidden layers in task 1 are updated by fine-tuning. Lines 5-7 carry out the update of weights and nodes in task 2 (by means of fine-tuning). Line 8 obtains the endmember and abundance matrices. Line 9 generates the interaction abundance and outlier matrices.

Algorithm 1: Pseudocode of the proposed algorithm

Input: Hyperspectral data $\mathbf{Y}$, Layer number m.
Output: Endmembers $\mathbf{E}$, Abundances $\mathbf{A}$, Virtual Abundance $\mathbf{B}$, Bilinear Component $\widehat{\mathbf{Y}}_N$.
for all layers **do**
/* *Initialization* */
1. Run VCA and FCLS for the initialization of task 1.
2. Initialize task 2 by using the factors resulting from task 1.
end for
repeat
/* *Task 1 (for linear components)* */
for all layers **do**
3. Update $\mathbf{E}_1$ and $\mathbf{A}_1$ in Eq. (15).
while reach the threshold of Armijo rule **do**
4. Update $\mathbf{E}_k$ and $\mathbf{A}_k$ in Eq. (19).
end while
end for
/* *Task 2 (for bilinear components)* */
for all layers **do**
5. Update $\mathbf{D}_r$ in Eq. (23).
6. Update $\mathbf{B}_r$ in Eq. (22).
7. Update $\mathbf{B}_m$ in Eq. (24).
end for
8. Obtain $\mathbf{E}$ and $\mathbf{A}$ in Eq. (20).
9. Obtain $\mathbf{B}$ and $\widehat{\mathbf{Y}}_N$ in Eq. (26).
until RE reaches convergence

IV. Experimental Results and Analysis

The effectiveness of the proposed method called hereinafter Deep Multi-task Bilinear Unmixing (DMBU) is evaluated by using both synthetic and real hyperspectral datasets. In this paper, four metrics are adopted for the assessment of results, including the spectral angle distance (SAD), root mean square error (RMSE), RE, and variance inflation factor (VIF).

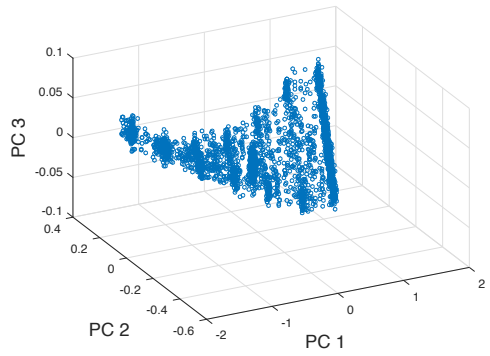


Fig. 2. Scatterplot of one of our simulated data sets, with 3364 pixels and SNR of 40 dB.

The SAD is used to measure the quality of the estimated endmember, which is defined as

$$\mathrm{SAD}(\mathbf{e}, \widehat{\mathbf{e}}) = \arccos\left(\frac{[\mathbf{e}, \widehat{\mathbf{e}}]}{\|\mathbf{e}\| \cdot \|\widehat{\mathbf{e}}\|}\right), \tag{27}$$

where $\widehat{\mathbf{e}}$ and $\mathbf{e}$ denote the estimated endmember and library spectrum, respectively, and the SAD is specified in radians.

The quality of the estimated abundances is measured by using the RMSE defined as

$$\mathrm{RMSE}(\widehat{\mathbf{a}}_g,\ \mathbf{a}_g) = \frac{1}{n}\sum_{g=1}^{n}\sqrt{\|\ \mathbf{a}_g - \widehat{\mathbf{a}}_g\|_2^2}, \tag{28}$$

where $\widehat{\mathbf{a}}_g$ and $\mathbf{a}_g$ are the corresponding estimated and actual abundance fractions. The RE was given in (10). Note that the RE should not be directly regarded as a metric for measuring the accuracy of unmixing results. Instead, it is used as a complementary evaluation, following [40], [64].

Moreover, we use the VIF to measure the degree of collinearity, according to the relevant works in [64]–[66]. This metric is defined as

$$\mathrm{VIF}_i = \frac{1}{1 - X_i^2}, \tag{29}$$

where X_i^2 denotes the multiple correlation between the i-th variable (an endmember, or a virtual endmember) and other explanatory variables. A detailed explanation of the VIF is available in [65]. Note that, in this paper, the VIF denotes the mean of all VIF_i.

The reminder of this section is organized as follows. Subsection IV-A describes the experiments with synthetic data sets. These experiments focus on the network parameters, the collinearity impact, and a comparative algorithm assessment. In subsection IV-B, some real hyperspectral images are used to further evaluate the effectiveness of our newly proposed DMBU. Note that all experiments have been performed in a workstation with Intel Core I7 CPU and 16 GB of RAM.

A. Experiments with Synthetic Data

The synthetic data set are generated using a bilinear mixture model (BMM). Three pure spectral signatures are randomly selected from the United States Geological Survey (USGS)

TABLE I
MEAN SADS (IN RADIANS), RMSES AND RES OBTAINED BY THE PROPOSED DMBU BY USING DIFFERENT BALANCE COEFFICIENTS, FOR A SYNTHETIC DATA SET WITH 3364 PIXELS AND SNR OF 40DB.

		SAD	RMSE	RE
$\mu = 0.9$	$\upsilon = 0.9$	0.0161	0.4208	0.0148
	$\upsilon = 0.5$	0.0149	0.4004	0.0127
	$\upsilon = 0.1$	0.0127	0.3813	0.0103
	$\upsilon = 0.05$	0.0130	0.3906	0.0105
$\mu = 0.5$	$\upsilon = 0.9$	0.0154	0.4255	0.0119
	$\upsilon = 0.5$	0.0144	0.4192	0.0114
	$\upsilon = 0.1$	0.0124	0.3702	0.0091
	$\upsilon = 0.05$	0.0129	0.3916	0.0092
$\mu = 0.1$	$\upsilon = 0.9$	0.0168	0.4569	0.0175
	$\upsilon = 0.5$	0.0135	0.4371	0.0087
	$\upsilon = 0.1$	0.0149	0.4424	0.0092
	$\upsilon = 0.05$	0.0184	0.4406	0.0118
$\mu = 0.05$	$\upsilon = 0.9$	0.0134	0.4221	0.0094
	$\upsilon = 0.5$	0.0141	0.4407	0.0101
	$\upsilon = 0.1$	0.0137	0.4445	0.0090
	$\upsilon = 0.05$	0.0152	0.4373	0.0086

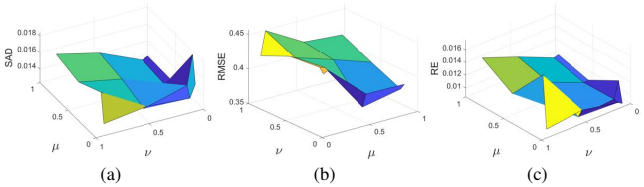


Fig. 3. Results of the proposed DMBU with respect to balance coefficients μ and υ. (a) SAD. (b) RMSE. (c) RE.

spectral library [1]. Each pixel in the synthetic data set has 224 spectral bands, covering the spectral range from 0.4 μm to 2.5 μm. The abundance settings follow a random distribution, and the maximum abundance purity of the synthetic data is set to 0.9, i.e., all pure pixels are removed. In our synthetic data experiments, we simulated several different scenarios: different numbers of pixels and different signal-noise ratio (SNR) levels. An illustrative scatterplot is given in Fig. 2, where the original data are projected onto the first three principal components (PCs) to facilitate visualization.

1) Parameter Analysis: In this test, the performance of network parameters is verified under different settings. The experimental results with respect to the coefficients in Eq. (11) and Eq. (21) are displayed in Table I. For the data used in Table I, the number of pixels is 3364, and the SNR is 40dB. For the autoencoders, the number of hidden layers is set to $m = 3$. The obtained results indicate that the best performance is obtained within a relatively wide range of values, i.e., $0.1 < \mu < 0.9$, $0.05 < \upsilon < 0.9$. Moreover, these results indicate that the proposed approach is quite insensitive in the aforementioned range. Fig. 3 visualizes the quantitative results obtained with respect to two balance coefficients. In the subsequent experiments, the coefficients are set as $\mu = 0.5$ and $\upsilon = 0.1$.

[1]https://speclab.cr.usgs.gov/spectrallib.html

TABLE II
MEAN SADS (IN RADIANS), RMSES, RES AND COMPUTATION TIME (IN SECONDS) OBTAINED BY THE PROPOSED DMBU WITH DIFFERENT m.

	m=1	m=2	m=3	m=4	m=5	m=6
SAD	0.0118	0.0114	0.0109	0.0107	0.0112	0.0110
RMSE	0.4030	0.3107	0.2662	0.2649	0.2719	0.2646
RE	0.0145	0.0135	0.0129	0.0128	0.0131	0.0127
Time	21.6308	34.9588	62.8081	85.2175	104.0968	137.3905

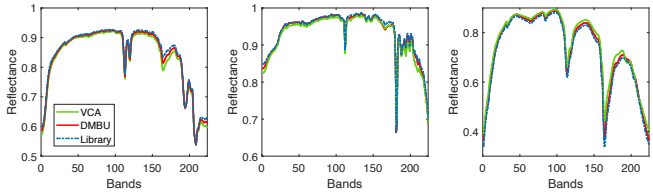


Fig. 4. Comparison between the estimated endmembers, along with the corresponding library signatures.

For autoencoder, the number of layers is also an important factor defining unmixing performance. Table II shows the differences in the obtained results when we assign a different number of layers in the proposed approach. In this test, the number of pixels is set to 3364, with SNR of 40dB. From Table II, we can observe that the results are better when m is set to 3 or 4. However, the performances are degraded when the parameter m is more than 4. In addition, more hidden layers means more computation time to implement the proposed approach. Considering the performance results and computation time in Table II, we set $m = 3$ in the subsequent experiments.

In order to graphically illustrate the effectiveness of the proposed method, the estimated endmembers via DMBU are shown in Fig. 4. The results obtained by the VCA are also displayed in Fig. 4, considering that they are used to initialize the proposed approach. In this test, the data includes again 3364 pixels and SNR of 40dB. Fig. 4 shows a graphical comparison between the reference library signatures and the estimated endmembers. Fig. 4 reveals that the endmembers estimated by DMBU are quite similar to the corresponding ones in the library.

Moreover, the proposed DMBU can also estimate abundance fractions. In this test, the number of pixels in the synthetic data is set to 100. The estimated abundances are illustrated in Fig. 5 (a). As it can be seen in Fig. 5 (a), the estimated abundances exhibit a very good match with regards to the ground-truth fractions. Fig. 5 (b) shows the original data $\mathbf{Y}$ and the reconstructed data $\widehat{\mathbf{Y}}$. The results in Fig. 5 (b) also indicate that the proposed DMBU can lead to very good reconstruction results.

2) Impact of Collinearity Analysis: In multivariate regression, collinearity is a common phenomenon whenever two or more of the variables in a model are moderately or highly correlated [67]. To our knowledge, SU can be regarded as a special optimization problem, so it is hard to absolutely avoid the influence of the collinearity among variables. For SU, the collinearity means that there are correlations among the endmembers, and the adverse impact of the correlations may

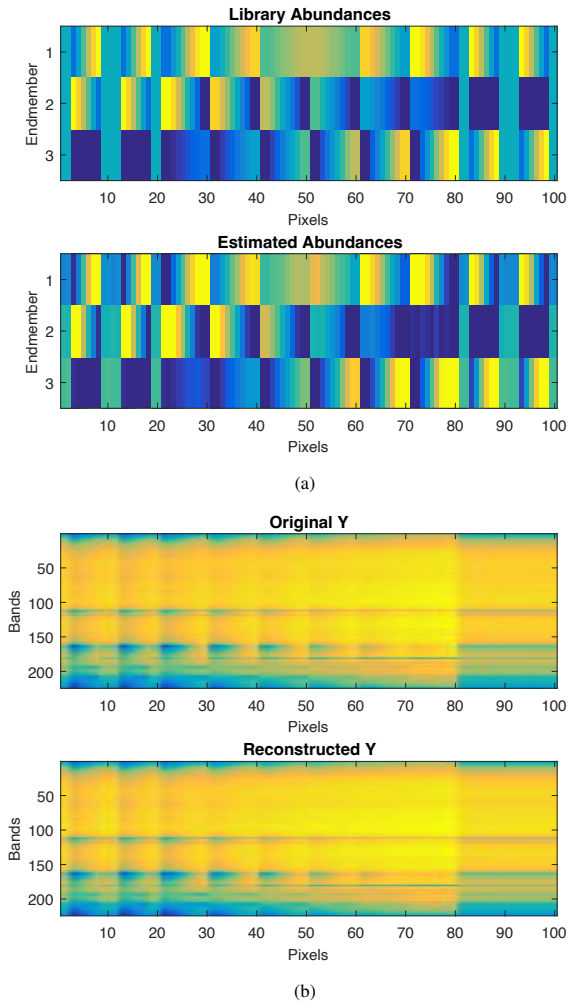


Fig. 5. (a) Top: ground-truth abundance. Bottom: abundances estimated by the DMBU. (b) Top: original data. Bottom: reconstructed data by DMBU. Bright yellow corresponds to large value. Conversely, dark blue corresponds to small value.

make regression sensitive to noise and cause a degradation of unmixing performance [64]–[66]. Therefore, it is necessary to evaluate the collinearity effect for virtual endmembers. In this test, we use the VIF to measure the degree of collinearity, under several different scenarios. Table III illustrates the VIFs obtained by the proposed DMBU. The results in Table III indicate that the VIFs of endmembers are negligibly small, while the virtual endmembers exhibit high collinearity. The reason is that the collinearity among independent variables is generally not significant. Geometrically speaking, a p-simplex is assumed from a p-dimensional polytope which is the convex hull of its $p+1$ vertices, and these vertices are mutually independent [68], [69]. For a hyperspectral image, the endmembers are located in the vertices of the simplex under the LMM, so the impact of collinearity is generally not significant. However, the virtual endmembers themselves are correlated, because they are produced by the Hadamard product between endmembers. Moreover, Table III also reveals that the collinearity effect decreases with the increase of the noise level.

TABLE III
VIFs OBTAINED BY DMBU IN SEVERAL DIFFERENT SCENARIOS.

VIF	100 Pixels		3364 Pixels	
	True Endmembers	Virtual Endmembers	True Endmembers	Virtual Endmembers
SNR=80dB	8.6352	92.2544	9.2458	97.1803
SNR=40dB	8.9893	109.0075	10.9495	89.6224
SNR=30dB	11.6849	193.0777	12.6841	204.4147
SNR=20dB	13.2706	422.2511	14.1632	551.9053

3) Comparison with Other Algorithms: In this section, our proposed DMBU is evaluated by comparing it with several widely used unmixing algorithms, such as N-FINDR [12], VCA [13], FCLS [54], $L_{1/2}$-NMF [20], GBM [39], Semi-NMF [40], and rNMF [41]. These unmixing algorithms were chosen as a comparative basis for the following reasons. In the selected methods, N-FINDR, VCA and FCLS are generally used for the initialization of other methods. $L_{1/2}$-NMF also adopts MUR to update variables. N-FINDR and VCA are used for extracting endmember signatures, while FCLS, GBM, and Semi-NMF can only used for obtaining abundance fractions. In the aforementioned approaches, $L_{1/2}$-NMF, rNMF, and DMBU are unsupervised algorithms which are able to estimate endmember signatures and abundance fractions, simultaneously. VCA, N-FINDR, FCLS, and $L_{1/2}$-NMF are approaches for linear unmixing. On the contrary, GBM, Semi-NMF, rNMF, and DMBU are nonlinear unmixing approaches. According to the overviews in [8], [32], the proposed DMBU is a nonlinear unmixing approach because it is based on a BMM.

In this comparison experiment, VCA is used to extract endmembers signatures before running FCLS, GBM, and Semi-NMF. Note that, for linear unmixing algorithms, the image reconstruction only relates to endmember signatures and abundance fractions. The quantitative results are shown in Table IV, and the results are obtained by the averaging of the results obtained after 10 independent Monte Carlo runs. From the results reported in Table IV, it can be observed that the proposed DMBU obtains competitive results when compared with other methods. More specifically, the proposed approach is able to obtain good results on the estimation of endmembers and abundances. This is demonstrated by the fact that it achieves better results in terms of SAD, RMSE and RE. It is noticeable that the advantages of our approach are suitable for the scenarios in which SNR level is more than 20 dB. The proposed approach can efficiently address the problem of mixed pixels in most real hyperspectral image scenes, as long as the SNR levels of these images are higher than 20 dB.

B. Experiments with Real Data

In this subsection, the proposed DMBU is applied to four real hyperspectral data sets: Moffett Field [70], Jasper Ridge [71], Urban [72], and Henry Island [73]. The first three data sets have been widely used for testing the performance of unmixing algorithms. The Henry Island data set is also used to conduct sub-pixel analysis for ecosystems research on mangrove forests. In these experiments, the parameters of

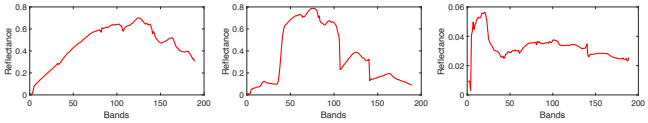
TABLE IV

MEAN SADS (IN RADIANS), RES AND RMSES, ALONG WITH THE STANDARD DEVIATIONS OBTAINED FROM 10 INDEPENDENT MONTE CARLO RUNS FOR THE CONSIDERED SYNTHETIC DATA. THE BEST RESULTS ARE DISPLAYED IN BOLD.

		3364 Pixels						
		N-FINDR+FCLS	VCA+FCLS	$L_{1/2}$-NMF	VCA+GBM	VCA+Semi-NMF	rNMF	DMBU
SNR=30 dB	SAD	0.0141±0.02%	0.0099±0.05%	0.0302±6.19%	0.0099±0.05%	0.0099±0.05%	**0.0091**±0.12%	0.0096±0.08%
	RMSE	0.3894±0.87%	0.3726±0.28%	0.9487±2.52%	0.3790±0.90%	0.3712±0.35%	0.3670±0.26%	**0.3651**±0.51%
	RE	0.0137±0.01%	0.0141±0.01%	0.0215±0.34%	0.0132±0.04%	0.0143±0.08%	0.0675±0.06%	**0.0116**±0.10%
SNR=20 dB	SAD	0.0376±0.16%	0.0176±0.42%	0.0394±2.92%	0.0176±0.42%	0.0176±0.42%	0.0405±0.56%	**0.0127**±0.73%
	RMSE	0.2201±0.62%	0.1603±0.58%	0.7257±5.32%	0.1625±2.05%	**0.1525**±0.45%	0.1610±0.81%	0.1626±0.93%
	RE	0.0934±0.19%	0.0873±0.25%	0.0856±0.53%	0.0994±0.37%	0.0977±0.30%	0.1547±0.55%	**0.0840**±0.42%
		676 Pixels						
		N-FINDR+FCLS	VCA+FCLS	$L_{1/2}$-NMF	VCA+GBM	VCA+Semi-NMF	rNMF	DMBU
SNR=30 dB	SAD	0.0162±0.04%	0.0129±0.03%	0.0328±0.85%	0.0129±0.03%	0.0129±0.03%	0.0098±0.09%	**0.0113**±0.08%
	RMSE	0.4005±0.21%	0.3334±0.83%	0.9828±1.96%	0.4018±1.05%	**0.3157**±0.45%	0.3420±0.62%	0.3493±0.65%
	RE	0.0183±0.02%	0.0154±0.06%	0.0201±1.01%	0.0157±0.04%	0.0136±0.07%	0.0680±0.08%	**0.0140**±0.07%
SNR=20 dB	SAD	0.0410±0.10%	0.0197±0.15%	0.0435±1.30%	0.0197±0.24%	0.0197±0.20%	0.0428±0.73%	**0.0184**±0.52%
	RMSE	0.2335±0.58%	0.2416±0.23%	0.8876±2.09%	0.4469±1.27%	0.2514±0.74%	**0.2402**±0.91%	0.3046±1.08%
	RE	0.0997±0.21%	0.0856±0.16%	0.0883±0.92%	0.0939±0.45%	0.0998±0.70%	0.1548±0.68%	**0.0839**±0.39%

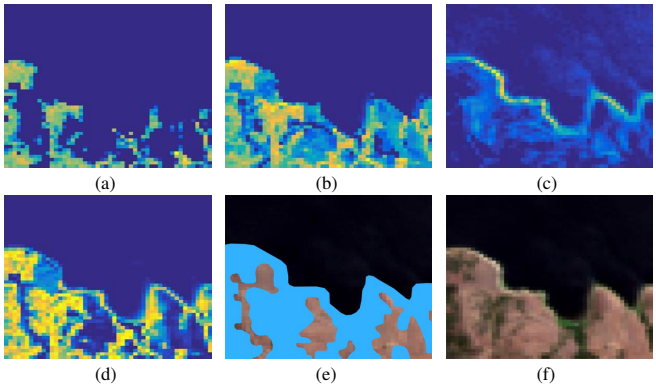
(a) (b) (c)

Fig. 6. Interaction abundance maps estimated by the DMBU in the Moffett Field data. (a) Soil-Vegetation. (b) Vegetation-Water. (c) Soil-Water.



(a) (b) (c)

Fig. 7. Endmembers estimated by the proposed DMBU in the Moffett Field dataset. (a) Soil. (b) Vegetation. (c) Water.

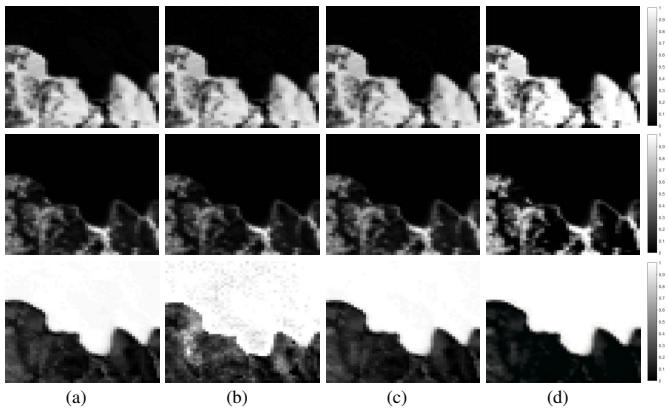


(a) (b) (c) (d) (e) (f)

Fig. 9. Interaction distributions estimated by different methods in the Moffett Field image. Bright yellow pixels correspond to large values. Conversely, dark blue pixels correspond to small ones. (a) GBM. (b) Semi-NMF. (c) rNMF. (d) DMBU. (e) Boundaries and mixing regions between materials, corresponding to the light blue area. (f) Real Moffett Field image.

DMBU follow the same settings established in the synthetic data experiments, i.e., $\mu = 0.5$, $\upsilon = 0.1$, and $m = 3$.

1) Experiment with Moffett Field Data Set: The Moffett Field data is a real hyperspectral image acquired over Moffett Field, CA, USA, in 1997. It was gathered by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) [70], [74].



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

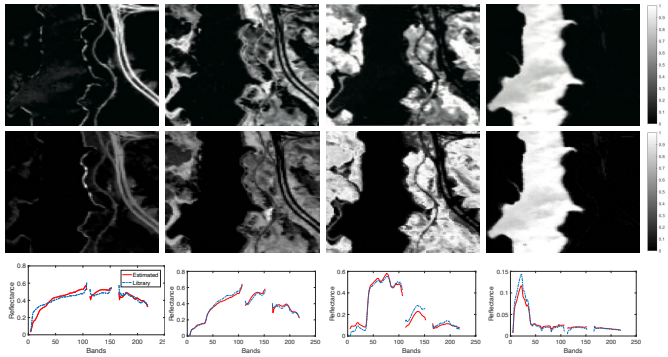
Fig. 8. Abundance maps estimated by different methods in the Moffett Field data. Top to bottom: Soil, Vegetation, Water. (a) GBM. (b) Semi-NMF. (c) rNMF. (d) DMBU.

In this experiment, the selected sub-scene comprises 50×50 pixels, with 189 bands covering the wavelength region from 0.4μm to 2.5μm (low SNR and water absorption bands were removed from the data). The selected region contains three endmembers, which are characteristic of the coastal area in the image: vegetation, water, and soil [70]. The sub-scene is widely used to evaluate algorithms for nonlinear unmixing in the remote sensing community [39].

Due to the fact that Moffett data lacks reference library signatures and truth abundances, our experiment just uses the RE to evaluate the accuracy of the obtained unmixing results. As shown in Table V, the proposed DMBU obtained the best RE value, although it takes longer computation time. Moreover, the result also reveals that linear unmixing approaches generally exhibit lower computational complexity than nonlinear ones. For illustrative purposes, Fig. 6 displays the interaction abundance maps obtained by DMBU. Specifically, the scattering interaction between water and soil is mainly distributed in the boundaries between regions, and the scattering between soil and vegetation is quite high. Fig. 7 shows the endmembers estimated by DMBU, and Fig. 8 shows an abundance comparison among the nonlinear unmixing ap-

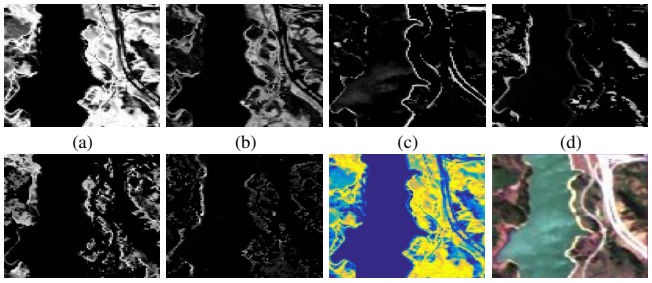
TABLE V
REs AND COMPUTATION TIME (IN SECONDS) OBTAINED BY DIFFERENT METHODS FOR THE MOFFETT FIELD DATA, WHERE THE BEST RESULTS ARE DISPLAYED IN BOLD.

Attribute	LSMM-based			NSMM-based			
Method	N-FINDR+FCLS	VCA+FCLS	$L_{1/2}$-NMF	VCA+GBM	VCA+Semi-NMF	rNMF	DMBU
RE	0.0210	0.0202	0.0280	0.0192	0.0223	0.0207	**0.0187**
Time	0.9372	**0.3282**	0.3526	123.1092	30.7614	40.3109	64.3542



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

Fig. 10. Results obtained by DMBU on the Jasper Ridge data. Top: reference abundance maps. Middle: estimated abundance maps. Bottom: comparison of the estimated endmembers (red curves) with the reference signatures (blue curves). (a) Road. (b) Soil. (c) Tree. (d) Water.



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

Fig. 11. Interaction abundance maps and interaction distribution map estimated by DMBU in the Jasper Ridge sub-image. (a) Soil-Tree. (b) Soil-Water. (c) Soil-Road. (d) Tree-Road. (e) Tree-Water. (f) Water-Road. (g) Interaction distribution map. (h) Real image.

proaches. In Fig. 8, DMBU obtained competitive results when compared with other approaches. From this figure, we can conclude that the proposed DMBU can estimate the abundance distribution of the water better than the other tested methods.

Fig. 9 illustrates the scattering interactions estimated by different nonlinear approaches. The interaction distribution maps of rNMF and DMBU are obtained from the interaction outliers, while the ones of GBM and Semi-NMF are obtained by multiplying the virtual endmembers and interaction abundances. To our knowledge, second-order scattering should be mainly present in the boundaries and mixed regions between materials. Compared with the real distribution of materials in Fig. 9, we can observe that the scattering effects predicted by the proposed DMBU are reasonable, i.e, the interactions are mainly distributed in the boundaries and mixed regions between the materials. However, the other tested methods seem to find some problems. These results are consistent with our introspections on the expected distribution of nonlinear interactions.

2) Experiment with Jasper Ridge Data Set: The Jasper Ridge data (512 $\times$ 614 pixels) was gathered by AVIRIS over Jasper Ridge in California. The data consists of 224 bands over the wavelength region from 0.38 μm to 2.5 μm. Considering water vapor and atmospheric effects, bands $1-3$, $108-112$, $154-166$, and $220-224$ were removed, leaving a total of 198 spectral bands for our experiments [71]. To reduce the complexity of the data, a sub-image with 100 $\times$ 100 pixels is selected from the original data set (this is a widely used image for testing unmixing algorithms). The sub-image contains 4 endmembers: road, soil, tree, and water [71].

Table VI indicates that the proposed DMBU obtained the best RMSE and RE scores in this experiment. Moreover, the quantitative results illustrate that the proposed DMBU obtained better SADs. For illustrative purposes, the endmember signatures and abundance maps are shown in Fig. 10, respectively. The middle part of Fig. 10 shows that the abundances estimated via DMBU, and it can be seen that these abundances match well with the corresponding references. The bottom of Fig. 10 reveals that the endmembers estimated via DMBU also exhibit a good match with the corresponding library signatures. The results in Table VI and Fig. 10, further demonstrate the effectiveness of the linear part of the proposed DMBU. On the other hand, Figs. 11 (a)-(f) illustrate the interaction abundance maps estimated by DMBU in the Jasper Ridge data. Fig 11 (g) shows the interaction outliers obtained via the proposed approach. Compared with the real color Jasper Ridge data displayed in Fig. 11 (h), most of the obtained results are reasonable, since the estimated interactions are distributed in the boundary and mixed regions between the constituent materials in the scene.

3) Experiment with Urban Data Set: The Urban data was acquired by the Hyperspectral Data Image Collection Experiment (HYDICE) instrument over the Copperas Cove, TX, US, in 1995 [2]. This region of the image contains a mixture of man-made objects and forestry. In the Urban data, the top of the scene contains a highway that crosses the region from left to right, a shopping mall along the highway, and a parking lot in front of the mall. The Urban data comprises a total of 307 $\times$ 307 pixels and 210 bands, with spectral resolution of 10 nm and spatial resolution of $2 \times 2\text{m}^2$ per-pixel, covering the wavelength region from 0.4μm to 2.5μm [4], [72]. Prior to the analysis, bands $1-4$, 76, 87, $101-111$, $136-153$ and $198-210$ were removed due to water absorption and

[2]https://www.agc.army.mil/Hypercube.

TABLE VI
SADs (IN RADIANS) RMSEs, AND REs OBTAINED BY DIFFERENT METHODS IN THE JASPER RIDGE DATA, WHERE THE BEST RESULTS ARE DISPLAYED IN BOLD.

Method		N-FINDR+FCLS	VCA+FCLS	$L_{1/2}$-NMF	VCA+GBM	VCA+Semi-NMF	rNMF	DMBU
SAD	Road	0.0906	0.1009	0.2560	0.1009	0.1009	**0.0983**	0.0991
	Soil	0.2254	0.2093	**0.1908**	0.2093	0.2093	0.1927	0.1920
	Tree	**0.1558**	0.1799	0.1844	0.1799	0.1799	0.1730	0.1752
	Water	0.1337	0.1324	0.1584	0.1324	0.1324	0.1326	**0.1319**
Mean SAD		0.1514	0.1556	0.1974	0.1556	0.1556	**0.1492**	0.1496
RMSE		0.2754	0.2733	0.2641	0.2736	0.2508	0.2479	**0.2474**
RE		0.0168	0.0170	0.0265	0.0191	0.0196	0.0201	**0.0162**

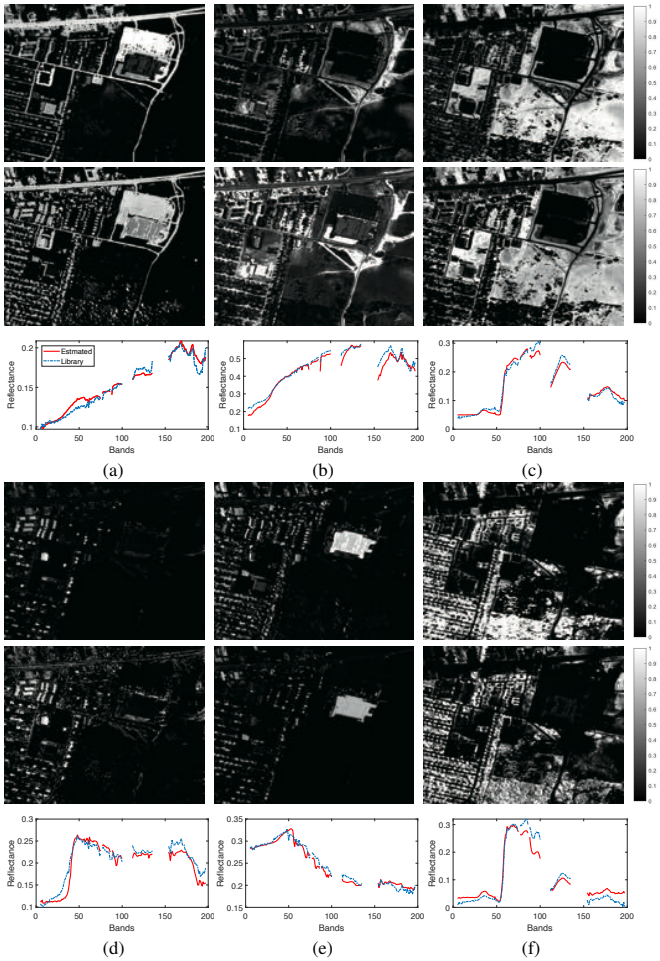


Fig. 12. Results estimated by the proposed DMBU in the Urban data. Top: reference abundance map. Middle: estimated abundance maps. Bottom: comparison of the estimated endmembers (red curves) with the reference signatures (blue curves). (a) Asphalt. (b) Dirt. (c) Grass. (d) Metal. (e) Roof. (f) Tree.

atmospheric effects in those bands, leaving a total of 162 spectral bands. In this experiment, the Urban data are used to further evaluate the proposed approach, and note that the scene includes six endmembers: asphalt, dirt, grass, metal, roof, and tree.

The quantitative results reported in Table VII reveal that the proposed DMBU obtained better results when compared with other methods in terms of the RMSE and RE values. For illustrative purposes, the estimated endmember signatures and

Fig. 13. (a) Real Urban image. (b) Interaction distribution map estimated by the proposed DMBU.

abundances are shown in Fig. 12. In Fig. 12, we can observe that the abundance fractions and endmember signatures estimated by the proposed approach provide a very good match with regards to the corresponding ones in the ground-truth. Fig. 13 displays the distribution map of interaction outliers which estimated by the DMBU. The figure reveals that the scattering interactions estimated by DMBU are reasonable, since they are distributed in the the boundary regions between the constituent materials in the scene.

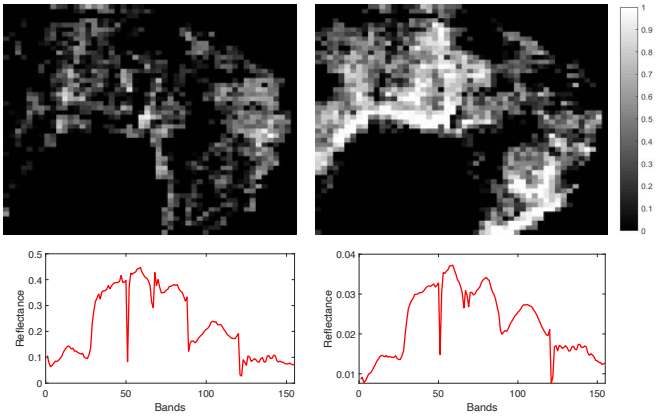
4) Experiment with Henry Island Data Set: The Henry Island data set was gathered by Earth-Observing-1 (EO-1) Hyperion and contains 137×187 pixels and 155 bands, with spatial resolution of $30m$. These data were obtained from the USGS Earth Resources Observation and Science Center through a data acquisition request to the satellite data provider. According to the description in [73], atmospheric correction has converted the data to reflectance units by using the FLAASH model available in ENVI software. In this experiment, we select a sub-scene with 50×50 pixels from the Henry Island data. According to [73], the area in the sub-scene contains two endmembers: 1) *Excoecaria Agallocha*, and 2) *Ceriops Decandra*.

The abundances and endmembers estimated by the proposed DMBU are depicted in Fig. 14. Fig. 15 shows the distribution of interaction abundances and outliers, reflecting the distribution range of scattering interactions on the data. According to [73], *Excoecaria Agallocha* and *Ceriops Decandra* grow together in the field. As a result, the interactive distribution in Fig. 14 is reasonable, as it can also be seen in the real image. In addition, after comparing the abundance map with the real image, we can observe that the results provided by our DMBU exhibit a good match with the distributions of the two plants in the real scene. These results further verify the effectiveness of our method.

TABLE VII

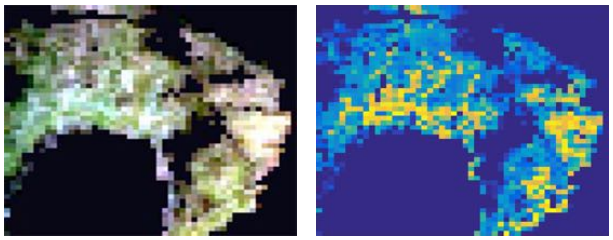
SADs (IN RADIANS) RMSEs, AND REs OBTAINED BY DIFFERENT METHODS FOR THE URBAN DATA, WHERE THE BEST RESULTS ARE DISPLAYED IN BOLD.

Method		N-FINDR+FCLS	VCA+FCLS	$L_{1/2}$-NMF	VCA+GBM	VCA+Semi-NMF	rNMF	DMBU
SAD	Asphalt	0.2174	0.0491	0.1505	0.0491	0.0491	0.0625	**0.0447**
	Dirt	0.1302	0.1089	0.1165	0.1089	0.1089	**0.0990**	0.0993
	Grass	**0.0691**	0.1251	0.2395	0.1251	0.1251	0.4164	0.1284
	Metal	0.8797	0.6921	**0.6807**	0.6921	0.6921	0.7038	0.6825
	Roof	0.3644	**0.1002**	0.1077	**0.1002**	**0.1002**	0.1763	0.1071
	Tree	0.4096	0.3547	0.3900	0.4096	0.4096	**0.1754**	0.2418
Mean SAD		0.3451	0.2384	0.2808	0.2384	0.2384	0.2722	**0.2173**
RMSE		0.3082	0.2294	0.2095	0.2070	0.2083	0.2134	**0.2062**
RE		0.0302	0.0226	0.0254	0.0219	0.0213	0.0206	**0.0201**



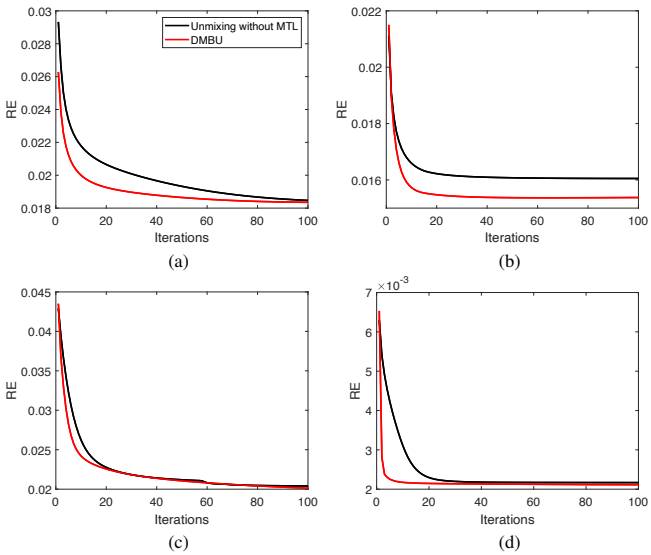
(a) (b)

Fig. 14. Abundance maps and endmember signatures estimated by the proposed DMBU in the Henry Island data: (a) *Excoecaria Agallocha*. (b) *Ceriops Decandra*.



(a) (b)

Fig. 15. (a) Real Henry Island image. (b) Interaction distribution map estimated by the proposed DMBU.



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

Fig. 16. Changes of the REs along with iterations, where these curves are obtained from different data sets. (a) Moffett Field. (b) Jasper Ridge. (c) Urban. (d) Henry Island.

5) Comparison on Optimization: The proposed DMBU adopts the MTL, aiming at handling the scattering interactions when estimating endmembers and abundances. It is worth noting, however, that the optimization performance may be different between the unmixing method with and without the MTL framework. For the proposed unmixing approach, the convergence on RE can reflect the optimization performance. The changes of the REs along with iterations are illustrated in Fig. 16, where the results are obtained on Moffett Field, Jasper Ridge, Urban, and Henry Island data sets, respectively. In Fig. 16, the red curves represent the results of DMBU, while the black ones are obtained for the proposed method without the MTL framework. To ensure a fair comparison, the number of iterations of the proposed methods are set to 100. From the curves shown in Fig. 16, we can observed that the optimizations are successful because these REs can reach convergence. Note that the unmixing method without the MTL framework only obtains endmembers and abundances, and the corresponding REs only relate to the linear components.

V. Conclusions and Future Lines

Autoencoders have successfully been used to solve unmixing problems by focusing on the linear mixture model. Moreover, most of the previous research works based on autoencoders only focused on increasing the depth of the networks to improve unmixing performance. However, our introspection in this paper is that simply increasing the depth of the networks may not be sufficient to model second-order scattering. To address this issue, in this paper we develop a new method that combines two deep autoencoders with a multi-task learning framework to conduct bilinear unmixing. Our framework, called DMBU, models the unmixing problem by using two different tasks that are fine-tuned by an iterative scheme. Our experimental results indicate that the newly developed DMBU is able to accurately estimate endmember signatures and abundance factions, while also modeling the impact of second-order scattering interactions.

As with any new approach, there are some unresolved issues that may present challenges over time. Specifically, our method is quite complex in computational terms. Moreover, the global optimization scheme adopted by our method represents a big challenge in practice. In order to enhance the practical exploitation of our proposed approach, we will develop parallel-optimized versions of the method and also consider in further detail the spatial correlation and spectral variability present in real scenarios.

APPENDIX

In this appendix, we provide the detailed derivation steps of the partial derivatives of $\mathcal{C}_{Task1}$. The partial derivatives of $\mathcal{C}_{Task1}$ with respect to $\mathbf{A}_k$ is given as

$$\begin{aligned}
&\nabla_{\mathbf{A}_k}\mathcal{C}_{Task\ 1}\\
&=\frac{1}{2}\frac{\partial\left\|\mathbf{Y}-\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_m\mathbf{A}_m\right\|_F^2}{\partial\mathbf{A}_k}+\frac{1}{2}\mu\frac{\partial\Re(\mathbf{A}_k)}{\partial\mathbf{A}_k}\\
&=\frac{1}{2}\partial\operatorname{Tr}(\mathbf{Y}^\top\mathbf{Y}-2\mathbf{Y}^\top\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_m\mathbf{A}_m+\mathbf{A}_m^\top\mathbf{E}_m^\top\mathbf{E}_{m-1}^\top\\
&\cdots\mathbf{E}_1^\top\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_m\mathbf{A}_m)/\partial\mathbf{A}_k+\frac{1}{2}\mu\frac{\partial\operatorname{Tr}\left(\mathbf{A}_k\mathbf{L}\mathbf{A}_k^\top\right)}{\partial\mathbf{A}_k}\\
&=\frac{1}{2}\frac{\partial\operatorname{Tr}\left(\mathbf{Y}^\top\mathbf{Y}-2\mathbf{Y}^\top\mathbf{\Psi}\mathbf{E}_k\mathbf{A}_k+(\mathbf{\Psi}\mathbf{E}_k\mathbf{A}_k)^\top\mathbf{\Psi}\mathbf{E}_k\mathbf{A}_k\right)}{\partial\mathbf{A}_k}\\
&+\frac{1}{2}\mu(\mathbf{A}_k\mathbf{L}+\mathbf{A}_k\mathbf{L}^\top)\\
&=\mathbf{E}_k^\top\mathbf{\Psi}^\top\mathbf{\Psi}\mathbf{E}_k\mathbf{A}_k-\mathbf{E}_k^\top\mathbf{\Psi}^\top\mathbf{Y}+\frac{1}{2}\mu(\mathbf{L}+\mathbf{L}^\top)\mathbf{A}_k\\
&=\mathbf{E}_k^\top\mathbf{\Psi}^\top(\mathbf{\Psi}\mathbf{E}_k\mathbf{A}_k-\mathbf{Y})+\mu\mathbf{A}_k\mathbf{L},
\end{aligned}\tag{30}$$

Similarly, the partial derivative of $\mathcal{C}_{Task1}$ with respect to $\mathbf{E}_k$ is obtained as

$$\begin{aligned}
&\nabla_{\mathbf{E}_k}\mathcal{C}_{Task\ 1}\\
&=\frac{1}{2}\frac{\partial\operatorname{Tr}\left(\mathbf{Y}^\top\mathbf{Y}-2\mathbf{Y}^\top\mathbf{\Psi}\mathbf{E}_k\widetilde{\mathbf{A}}_k+(\mathbf{\Psi}\mathbf{E}_k\widetilde{\mathbf{A}}_k)^\top\mathbf{\Psi}\mathbf{E}_k\widetilde{\mathbf{A}}_k\right)}{\partial\mathbf{E}_k}\\
&=\mathbf{\Psi}^\top\mathbf{\Psi}\mathbf{E}_k\widetilde{\mathbf{A}}_k\widetilde{\mathbf{A}}_k^\top-\mathbf{\Psi}^\top\mathbf{Y}\widetilde{\mathbf{A}}_k^\top\\
&=\mathbf{\Psi}^\top(\mathbf{\Psi}\mathbf{E}_k\widetilde{\mathbf{A}}_k-\mathbf{Y})\widetilde{\mathbf{A}}_k^\top.
\end{aligned}\tag{31}$$

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the Anonymous Reviewers for their constructive comments and suggestions, which significantly improved the quality of this manuscript.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] K. R. Thorp, A. N. French, and A. Rango, "Effect of image spatial and spectral characteristics on mapping semi-arid rangeland vegetation using multiple endmember spectral mixture analysis (mesma)," *Remote Sensing of Environment*, vol. 132, pp. 120–130, May. 2013.

[3] X. Jin and Y. Gu, "Superpixel-based intrinsic image decomposition of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4285–4295, Aug. 2017.

[4] X. Xu, J. Li, C. Wu, and A. Plaza, "Regional clustering-based spatial preprocessing for hyperspectral unmixing," *Remote Sensing of Environment*, vol. 204, pp. 333–346, Jan. 2018.

[5] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.

[6] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, Nov. 2019.

[7] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse nmf for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2815–2826, Oct. 2013.

[8] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[9] M. Tang, L. Gao, A. Marinoni, P. Gamba, and B. Zhang, "Integrating spatial information in the normalized p-linear algorithm for nonlinear hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1179–1190, Apr. 2018.

[10] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 240–251, Jan. 2016.

[11] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[12] M. E Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proceedings of SPIE Image Symposium V*, 1999, vol. 3753, pp. 266–277.

[13] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[14] A. Zare, P. Gader, O. Bchir, and H. Frigui, "Piecewise convex multiple-model endmember detection and spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2853–2862, Nov. 2013.

[15] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5067–5082, 2015.

[16] N. Dobigeon, S. Moussaoui, M. Coulon, J. Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[17] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4775–4789, Aug. 2016.

[18] D. Hong and X. X. Zhu, "Sulora: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1351–1363, Oct. 2018.

[19] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6076–6090, Jul. 2016.

[20] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[21] X. Feng, H. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, Oct. 2018.

[22] K. Janod, M. Morchid, R. Dufour, G. Linars, and R. De Mori, "Denoised bottleneck features from deep autoencoders for telephone conversation analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 9, pp. 1809–1820, Sep. 2017.

[23] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[24] A. Lemme, R. F. Reinhart, and J. J. Steil, "Online learning and generalization of parts-based image representations by non-negative sparse autoencoders," *Neural Networks*, vol. 33, no. 9, pp. 194–203, Sep. 2012.

[25] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *Proceedings of 7th Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS)*, 2015, pp. 1–4.

[26] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1427–1431, Sep. 2018.

[27] Y. Qu and H. Qi, "udas: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1698–1712, Mar. 2019.

[28] S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 482–496, Jan 2019.

[29] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep auto-encoder networks for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[30] B. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Hyperspectral unmixing using a neural network autoencoder," *IEEE Access*, vol. 6, pp. 25646–25656, Mar. 2018.

[31] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

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

[33] R. R. Close, P. D. Gader, and J. Wilson, "Hyperspectral unmixing using macroscopic and microscopic mixture models," *Journal of Applied Remote Sensing*, vol. 8, no. 1, pp. 1–16, Apr. 2014.

[34] B. Hapke, "Bidirectional reflectance spectroscopy: I. theory," *Journal of Geophysical Research Atmospheres*, vol. 86, no. B4, pp. 3039–3054, Apr. 1981.

[35] Y. Shkuratov, L. Starukhina, H. Hoffmann, and G. Arnold, "A model of spectral albedo of particulate surfaces: Implication to optical properties of the moon," *Icarus*, vol. 137, no. 2, pp. 235–246, Feb. 1999.

[36] B. T. Draine, "The discrete-dipole approximation and its application to interstellar graphite grains," *Astrophysical Journal*, vol. 333, pp. 848–872, Oct. 1988.

[37] B. Somers, K. Cools, S. Delalieux, J. Stuckens, D.V.Zande, W. W. Verstraeten, and P. Coppin, "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards," *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1183–1193, Jun. 2009.

[38] C. C. Borel and S. A. W. Gerstl, "Nonlinear spectral mixing model for vegetative and soil surfaces," *Remote Sensing of Environment*, vol. 47, no. 3, pp. 403–416, Mar. 1994.

[39] A. Halimi, Y. Altmann, N. Dobigeon, and J. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

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

[41] C. Fevotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.

[42] C. Zhao, G. Zhao, and X. Jia, "Hyperspectral image unmixing based on fast kernel archetypal analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 331–346, Jan. 2017.

[43] F. Zhu and P. Honeine, "Biobjective nonnegative matrix factorization: Linear versus kernel-based models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4012–4022, Jul. 2016.

[44] W. Luo, L. Gao, R. Zhang, A. Marinoni, and B. Zhang, "Bilinear normal mixing model for spectral unmixing," *IET Image Processing*, vol. 13, no. 2, pp. 344–354, Feb. 2019.

[45] Y. Su, J. Li, H. Qi, P. Gamba, A. Plaza, and J. Plaza, "Multi-task learning with low-rank matrix factorization for hyperspectral nonlinear unmixing," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 2127–2130.

[46] E. Theodoros and P. Massimiliano, "Regularized multi-task learning," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, vol. 9, pp. 109–117.

[47] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in *Proceedings of Advances in Neural Information Processing Systems*, 2006, pp. 41–48.

[48] C-. T. Lu, L. He, W. Shao, B. Cao, and P. S. Yu, "Multilinear factorization machines for multi-task multi-view learning," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. Apr. 2017, vol. 9, pp. 354–379, ACM.

[49] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, Dec. 2008.

[50] L. S. T. Ho, V. Dinh, and C. Nguyen, "Multi-task learning improves ancestral state reconstruction," *Theoretical Population Biology*, vol. 126, pp. 33–39, Apr. 2019.

[51] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Spectral-spatial hyperspectral unmixing using multitask learning," *IEEE Access*, vol. 7, pp. 148861–148872, 2019.

[52] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 417–429, Mar. 2017.

[53] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.

[54] D. C. Heinz and C. I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar.2001.

[55] A. Cichocki and R. Zdunek, "Multilayer nonnegative matrix factorization using projected gradient approaches," *International Journal of Neural Systems*, vol. 17, no. 6, pp. 431–446, Dec. 2007.

[56] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer nmf," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 38–42, Jan. 2015.

[57] Zuoyu Zhang, Shouyi Liao, Hexin Zhang, Shicheng Wang, and Yongchao Wang, "Bilateral filter regularized l2 sparse nonnegative matrix factorization for hyperspectral unmixing," *Remote Sensing*, vol. 10, no. 6, pp. 2072–4292, May. 2018.

[58] Y. Hao, C. Han, G. Shao, and T. Guo, "Generalized graph regularized non-negative matrix factorization for data representation," in *Proceedings of the 2012 International Conference on Information Technology and Software Engineerin*, 2013, pp. 1–12.

[59] O. Eches, N. Dobigeon, C. Mailhes, and J. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010.

[60] Daniel D. Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," *Proceedings of Advances in Neural Information Processing Systems 13*, pp. 556–562, 2001.

[61] J. Nocedal and S. J. Wright, "Numerical optimization," *New York, NY, Springer*, pp. 33–36, 2006.

[62] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, Jun. 2010.

[63] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust l_1-norm," in *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*.

[64] B. Yang, B. Wang, and Z. Wu, "Nonlinear hyperspectral unmixing based on geometric characteristics of bilinear mixture models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 694–714, Feb. 2018.

[65] X. Chen, J. Chen, X. Jia, B. Somers, J. Wu, and P. Coppin, "A quantitative analysis of virtual endmembers' increased impact on the collinearity effect in spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, pp. 2945–2956, Aug. 2011.

[66] L. Ma, J. Chen, Y. Zhou, and X. Chen, "Two-step constrained nonlinear spectral mixture analysis method for mitigating the collinearity effect,"

IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 5, pp. 2873–2886, May. 2016.

[67] M. H. Graham, "Confronting multicollinearity in ecological multiple regression," *Ecology*, vol. 84, no. 11, pp. 2809–2815, Nov. 2003.

[68] W. Rudin, *Principles of Mathematical Analysis (Third Edition)*, McGraw-Hill, New York, NY USA, 1976.

[69] R. Carter Hill and Lee C. Adkins, *A Companion to Theoretical Econometrics*, John Wiley and Sons, Ltd, Texas, TX, USA, 2007.

[70] E. Christophe, D. Leger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2103–2114, Sep. 2005.

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

[72] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 101–118, Feb. 2014.

[73] S. Chakravortty, J. Li, and A. Plaza, "A technique for subpixel analysis of dynamic mangrove ecosystems with time-series hyperspectral image data," *IEEE Journal of Selected Topics in Applied Earth*, vol. 11, no. 4, pp. 1244–1252, Apr. 2018.

[74] N. Dobigeon, J. Tourneret, and C. Chang, "Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.