



# Multiple-*Priors* Ensemble Constrained Nonnegative Matrix Factorization for Spectral Unmixing

Kewen Qu , *Student Member, IEEE*, and Wenxing Bao , *Member, IEEE*

**Abstract**—Nonnegative matrix factorization (NMF) is widely used in unmixing issue in recent years, because it can simultaneously estimate the endmembers and abundances. However, most existing NMF-based methods only consider single matrix constraints and the other one is ignored. In fact, due to the influence of various noise, the regularization effectiveness based on the single matrix constraint method may be limited. In addition, hyperspectral images contain a variety of *prior* information, while many approaches usually only consider one of the *priors*, and the synergistic effect of multiple *priors* unions and two matrix joint constraints is neglected. In this article, to overcome this limitation, we propose a new blind unmixing scheme, called multiple-*priors* ensemble constrained NMF. The article first analyses the HSI intrinsic feature *priors* from both geometric and statistical aspects, and three important *priors* learners are defined. Then, three learners are jointly introduced into the NMF model and work together for the first time to impose constraints on both the endmember and the abundance matrix. In order to effectively solve the proposed model, Barzilai–Borwein stepsize strategy accelerates optimization algorithm is developed by using the variable splitting and augmented Lagrangian framework. The effectiveness and superiority of the proposed method are demonstrated by comparing with other advanced approaches on both synthetic and real world datasets.

**Index Terms**—Barzilai–Borwein stepsize, hyperspectral unmixing, multiple-*priors*, nonnegative matrix factorization (NMF), variable splitting and augmented Lagrangian.

## I. INTRODUCTION

WITH the rapid development of hyperspectral imaging technology, the analysis and processing of hyperspectral remote sensing imagery (HSI) have attracted great attention, such as classification [1], subpixel mapping [2], target detection, and recognition [3], etc. Due to the complexity of the topography and the limitation of the spatial resolution of the sensors, a single pixel in HSI inevitably contains several kinds of materials [4]. Mixed pixels hinder the in-depth analysis and application of HSI and cause hyperspectral unmixing (HU) issues. Therefore, the task of HU is to decompose mixed pixels into a collection of spectral signatures (or endmembers) and their corresponding fractional abundances [5].

Manuscript received September 16, 2019; revised February 16, 2020; accepted February 21, 2020. Date of publication February 28, 2020; date of current version March 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61461003, and in part by the Image and Intelligent Information Processing Team and Science Foundation of North Minzu University, China under Grants PY1901 and 2016JSKY04. (Corresponding author: Wenxing Bao.)

The authors are with the School of Computer and Information, Hefei University of Technology, Hefei 230009, China, and also with the School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China (e-mail: kewen.qu@aliyun.com; bw71@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.2976602

For HU problem, a mass of approaches have been proposed based on linear mixing model (LMM) [5], which can be roughly divided into geometrical, statistical, sparse regression, and deep learning (DL) based methods. Specifically, the geometrical-based approaches assume that the vertices of a simplex formed by data points in the feature space correspond to endmembers, which mainly include the subspace projection methods [6], the maximum simplex [7], and the minimum simplex [8], [9] methods. Generally, the accuracy of abundance inversion of geometrical methods depends heavily on the quality of endmembers extraction. The DL approaches can capture the hierarchical structure of hidden information in hyperspectral data, which overcomes the limitations of single-layer information and further mine deep knowledge in the scene [10].

In the statistical framework, HU is formulated as a statistically independent or uncorrelated Bayesian inference problem, and well-known algorithms such as nonnegative matrix factorization (NMF) [11]. NMF has attracted a lot of attention in HU, as it can simultaneously estimate endmembers and their corresponding abundances without pure pixel assumption [12]–[15]. Unfortunately, there are a large number of local minima due to the nonconvexity of the NMF model. Therefore, additional constraints need to be incorporated to shrink the solution space.

In the last decades, various *priors* have been imposed on the endmember or abundance matrices by incorporating spectral, spatial, or joint spectral-spatial information into the NMF model. For the endmember matrix, the minimum simplex volume [12], the endmember minimum dispersion [16], and the endmember dissimilarity [17] are the common constraints. In [12], Miao and Qi first introduced the minimum volume regularization into the NMF framework and proposed a geometrical-statistical hybrid method. In addition, spectral variability is also introduced into the unmixing scenario as knowledge, and has received more attention in recent years [18]–[21]. In terms of abundance matrix, sparsity and smoothness are the two most important *priors*. Driven by this semantics, the  $L_{1/2}$  regularizer [22] is introduced to enforce sparsity of abundance and Qian *et al.* [13] proposed a sparsity constraint NMF. The  $L_{1/2}$  regularizer is explicitly compatible with the full additivity constraint of abundances. With the aim of exploring the latent manifold structure of the HSI, Lu *et al.* [14] proposed graph-regularized sparse NMF, which used the nearest neighbor graph to model the local information of high-dimensional space so that neighboring pixels can share similar abundances. Similarly, approaches include graph Laplacian [23] and hypergraph learning [24].

Total variation (TV) regularization is another popular spatial smoothing model in recent years, because the TV can preserve the image edge information and promote the piecewise smoothness of the abundance maps [15], [25]–[28]. In [15], a HTV (band-by-band TV) was embedded into a weighted  $L_1$ -norm sparse framework, and He *et al.* proposed a TV regularized reweighted sparse NMF. The TV was also introduced into non-negative tensor factorization (NTF) [25] and deep multilayer NMF [10].

Despite the commendable success, all of the aforementioned methods in HU issue, there is still room to further improve the unmixing performance: 1) The LMM can be described either from convex-geometric perspective or from statistical point of view. Therefore, hyperspectral data not only reflects the *prior* information of convex-geometric, but also contains statistical *prior* knowledge, which characterize the intrinsic properties of the hyperspectral data from different aspects. In order to model the unmixing issue more effectively, the above *priors* should be utilized at the same time. How to make full use of HSI various *priors* in the unified unmixing model lacks further research; 2) The NMF-based unmixing model includes two equally important matrix factors, i.e., endmember and abundance. However, most of the existing NMF-based methods only consider one matrix constraint and the other one is ignored. In addition, the observed hyperspectral data are inevitably destroyed by various noise due to the complicated transmission environment. Therefore, in the NMF-based unmixing model, the effectiveness of regularization based on the single matrix constraint method is limited. To further improve the accuracy of the unmixing, constraints should be imposed on both the endmember and the abundance matrix to more effectively regularize the nonconvex model and shrink the solution space.

In this article, to overcome this limitation, we propose a new blind unmixing method, called multiple-*priors* ensemble constrained NMF (MPEC-NMF). In hyperspectral geometric *priors*, simplex minimum volume (MV) is the most important *prior*. In statistical *priors*, abundance sparsity and abundance smoothness also received attention from a large number of researchers. However, the MV, sparsity and smoothness, which are important *priors* information, have not been deeply studied in the same unmixing model. The purpose of this method is to explore a new scheme, that is, by combining multiple *prior* information and imposing constraints on two matrices in the NMF model. The *ensemble* means the union of geometric and statistical *priors* and double matrix constraints. The experimental results demonstrate the effectiveness and superiority of the proposed method. This article is the continuation of the conference paper in [29].

The main contributions of this article are summarized as follows.

- 1) A new exploration scheme is proposed for the unmixing task, that is, by unifying geometric *priors* and statistical *priors* into the same unmixing model and imposing constraints on the endmember and the abundance matrix in the NMF model simultaneously, which overcomes the limitations of the traditional single matrix constraint method.

- 2) The minimum simplex, abundance sparsity, and abundance smoothness work together in the same model for the first time. Specifically, in our proposed model, the minimum simplex *prior* adopts the MV constraint, and the abundance sparsity and abundance smoothness *prior* adopt the reweighted  $L_{1/2}$  and TV operators, respectively. The contribution of each *prior* regularizer to the proposed model is analyzed and discussed by experiments.
- 3) We developed an optimization algorithm to solve the model by using the variable splitting and augmented Lagrangian. In order to improve the efficiency of the approach, the preconditioned conjugate gradient (PCG) method and the Barzilai–Borwein (BB) gradient method are used to embed the MPEC-NMF to accelerate the convergence of the algorithm, and the computational complexity of the algorithm is briefly analyzed.

The remainder of this article is organized as follows. Section II presents the LMM and basic NMF model. The proposed model and its corresponding optimization method are presented in Sections III and IV, respectively. In Section V, both synthetic data and real hyperspectral data experiments are described and analyzed, and the conclusions are drawn in Section VI.

## II. BACKGROUND

### A. Linear Mixing Model

Based on the LMM, the HSI dataset can be expressed as

$$\mathbf{Y} = \mathbf{E}\mathbf{A} + \mathbf{G} \quad (1)$$

where  $\mathbf{Y} \in \mathfrak{R}_+^{L \times N}$  refers to the observation HSI matrix, with  $L$  bands and  $N$  pixels.  $\mathbf{E} \in \mathfrak{R}_+^{L \times d}$  consists of the spectral signatures with  $d$  endmembers,  $\mathbf{A} \in \mathfrak{R}_+^{d \times N}$  represents the abundance matrix for all endmembers, and  $\mathbf{G} \in \mathfrak{R}^{L \times N}$  represents the additive noise matrix.

Two important constraints are usually adopted, namely abundance nonnegative constraint (ANC) and abundance sum-to-one constraint (ASC) as follows:

$$\mathbf{A} \succeq 0, \mathbf{1}_d^T \mathbf{A} = \mathbf{1}_N^T$$

where  $\mathbf{1}_d^T$  and  $\mathbf{1}_N^T$  represent all-one vectors with size  $d$  and size  $N$ , respectively. The symbol  $\succeq$  denotes componentwise inequality.

### B. Nonnegative Matrix Factorization

NMF can approximate a large nonnegative matrix into two low-rank nonnegative matrix factors [11]. Precisely, given an observation HSI data  $\mathbf{Y} \in \mathfrak{R}^{L \times N}$  and a positive number  $d < \min(L, N)$ . The purpose of NMF is to find two nonnegative matrices  $\mathbf{E} \in \mathfrak{R}^{L \times d}$  and  $\mathbf{A} \in \mathfrak{R}^{d \times N}$  by minimizing the following objective function:

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{A}} \quad & \frac{1}{2} \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \mathbf{E} \succeq 0, \mathbf{A} \succeq 0, \mathbf{1}_d^T \mathbf{A} = \mathbf{1}_N^T \end{aligned} \quad (2)$$

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

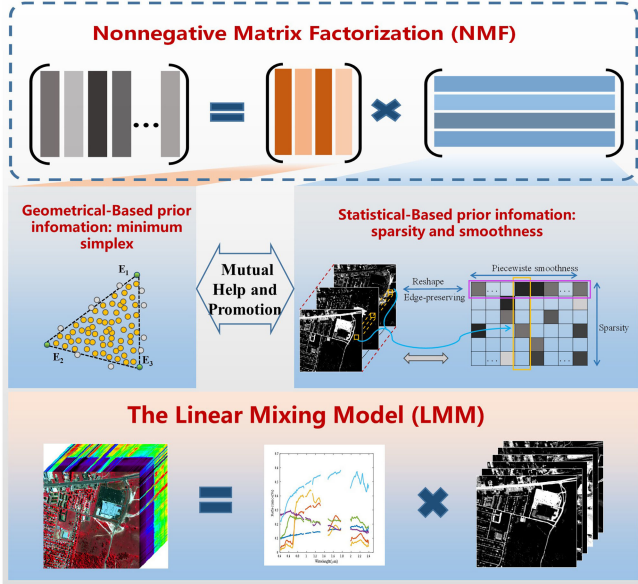


Fig. 1. Exhibits an illustration of the unmixing procedures for proposed method.

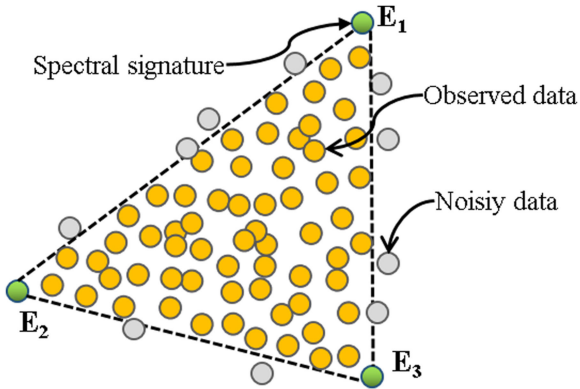


Fig. 2. Simplex illustrations: Three endmembers in 2-D space.

### III. PROPOSED MODEL

In this section, we present a new unmixing model, which unifies both geometric and statistical *prior* knowledge into the NMF framework, as shown in Fig. 1. Next, we first analyze the *priors* of HSI data, and then describe our proposed MPEC-NMF unmixing model in detail.

#### A. Geometrical-Based Prior Learner: Minimum Simplex

From the perspective of convex-geometry, the simplex volume defined by endmembers is the minimum of all the simplexes formed by the data cloud. The volume constraint cannot be only compact the simplex size, provide vertices that better approximate a given dataset, but also effectively cope with the effects of noisy pixels, and has a strong global constraint. As presented in Fig. 2, it can be seen that the observation data are contained in the simplex and the noise is excluded. Therefore, we imposed the MV constraint on the endmember matrix to improve the

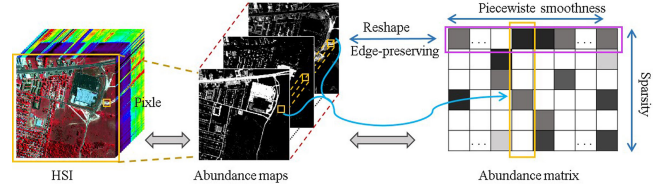


Fig. 3. Two *priors* of the abundance: The sparsity from column perspective and the piecewise smoothness from row perspective.

spectral signatures identification capability. The calculation of the volume of simplex benefited from Miao and Qi's outstanding work [12], which was presented as follows:

$$\phi_{\text{vol}}(\mathbf{E}) \propto \det^2(\mathbf{C} + \mathbf{B}\mathbf{U}^T(\mathbf{E} - \psi\mathbf{1}_d^T))/((d-1)! \quad (3)$$

where  $\mathbf{U} \in \mathfrak{R}^{L \times (d-1)}$  is a projection matrix defined as the  $d-1$  most significant principal components of the observed data  $\mathbf{Y}$  through the principal component analysis. The column vector  $\psi$  denotes the data mean,  $\mathbf{C} = [\mathbf{1}_d^T; \mathbf{0}]$  and  $\mathbf{B} = [\mathbf{0}_{d-1}^T; \mathbf{I}]$ ,  $\mathbf{1}$  and  $\mathbf{0}$  are vectors of all-ones and -zeros, respectively,  $\mathbf{0}$  is  $(d-1) \times d$  all-zero matrix, and  $\mathbf{I}$  is a  $(d-1) \times (d-1)$  identity matrix.

#### B. Statistical-Based Prior Learner: Sparsity and Smoothness

From a statistical perspective, the pixels in HSI are commonly a linear accumulation of a few signatures in the scene, and the column vectors of the abundance matrix should be sparse [13], as shown in Fig. 3. In addition, since the changes of ground materials distribution are generally slow and abrupt changes rarely appear, thus the abundance of two adjacent pixels corresponding to the same endmember should be similar, and abrupt changes occur only at the edge of the surface, hence the abundance maps should be piecewise smooth. Sparsity and smoothness reveal that the distribution characteristics and local affinity properties of earth's surface, which is an important *prior* knowledge of the hyperspectral data.

In recent years, compressed sensing [30] and regularizer theory [22], [31], [32] have shown that the  $L_{1/2}$  operator has unbiasedness and Oracle properties. The solution of the  $L_{1/2}$  operator is sparser than that of the  $L_1$  regularizer, while solving the  $L_{1/2}$  operator is much simpler than solving the  $L_0$  regularizer. Therefore, the  $L_{1/2}$  quasi-norm is a very popular sparsity-induced regularizer. In this article, the abundance sparsity *prior*-learner is defined by

$$\phi_{\text{spa}}(\mathbf{A}) = \|\mathbf{A}\|_{1/2} \quad (4)$$

where

$$\|\mathbf{A}\|_{1/2} = \sum_{i,j=1}^{d,N} \mathbf{A}_{ij}^{1/2}.$$

The  $L_{1/2}$  regularization, however, leads to a nonconvex, non-smooth, and non-Lipschitz optimization problem that is difficult to solve [22]. In order to solve the  $L_{1/2}$  optimization problem efficiently, which inspired by [33], a weighted method is used to convert  $L_{1/2}$  nonconvex problem into  $L_1$  convex optimization problem using a reweighted iterative strategy, and (4) becomes



the following form:

$$\phi_{\text{spa}}(\mathbf{A}) = \|\mathbf{W} \odot \mathbf{A}\|_1 \quad (5)$$

where  $\mathbf{W}$  is the weighted matrix with nonnegative elements, and operator  $\odot$  means elementwise multiplication. The weight matrix  $\mathbf{W}$  used for the next iteration is computed from the abundance matrix of the current solution

$$\mathbf{W}_{i,j}^{k+1} := 1 / \left( \sqrt{|\mathbf{A}_{i,j}^k|} + \epsilon \right) \quad (6)$$

where  $\mathbf{A}_{i,j}^k$  represents the abundance matrix of the  $k$ th iteration and  $\epsilon$  is any fixed positive real number.

In order to promote the piecewise smoothness of the abundance and preserve the spatial edge information, TV operator are adopted, which is defined as follows:

Let  $\nabla_h : \mathfrak{R}^{m \times n} \rightarrow \mathfrak{R}^{m \times n}$  denote a linear operator computing the horizontal differences, i.e.,  $\nabla_h \mathbf{A} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n]$ , where  $\mathbf{q}_i = \mathbf{a}_i - \mathbf{a}_{i_h}$ , with  $i$  and  $i_h$  denoting a pixel and its horizontal neighbor. Let  $\nabla_v : \mathfrak{R}^{m \times n} \rightarrow \mathfrak{R}^{m \times n}$  be defined in a similar way for the vertical differences, i.e.,  $\nabla_v \mathbf{A} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$ , where  $\mathbf{v}_i = \mathbf{a}_i - \mathbf{a}_{i_v}$ , with  $i$  and  $i_v$  denoting a pixel and its vertical neighbor. With these two difference operators, we define the TV

$$\nabla \mathbf{A}^T = \begin{pmatrix} \nabla_h \mathbf{A} \\ \nabla_v \mathbf{A} \end{pmatrix}$$

where  $\nabla = \begin{pmatrix} \nabla_h \\ \nabla_v \end{pmatrix}$  denotes TV operator. The symbol  $\nabla$  is applied on  $\mathbf{A}^T$  since each abundance map is along rows of the abundance matrix. The abundance smoothness *prior*-learner is defined as follows:

$$\phi_{\text{TV}}(\mathbf{A}) = \|\nabla \mathbf{A}^T\|_1. \quad (7)$$

### C. MPEC-NMF Model

By incorporating formulas (3), (5), and (7) into (2), the objective function of the MPEC-NMF model was obtained as follows:

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{A}} \quad & \frac{1}{2} \|\mathbf{Y} - \mathbf{EA}\|_F^2 + \lambda \phi_{\text{vol}}(\mathbf{E}) + \alpha \phi_{\text{TV}}(\mathbf{A}) + \beta \phi_{\text{spa}}(\mathbf{A}) \\ \text{s.t.} \quad & \mathbf{E} \succeq 0, \mathbf{A} \succeq 0, \mathbf{1}_d^T \mathbf{A} = \mathbf{1}_N^T \end{aligned} \quad (8)$$

where the first term is the fidelity, and the second to the fourth terms are three *prior*-learners. The parameters  $\lambda$ ,  $\alpha$ , and  $\beta$  control the contribution of the corresponding *priors*, respectively, which balance the accuracy of the reconstruction error.

In model (8), by synthesizing geometric and statistical *priors*, achieving multiple *priors* ensemble constraints on the endmember and the abundance matrices. Fig. 1 exhibits an illustration of the HU procedures. We use a variable splitting and augmented Lagrangian method [34] to solve the optimization model (8). Next, we will give detailed solving steps and algorithm flow.

## IV. OPTIMIZATION ALGORITHM

Clearly, the model (8) is a nonsmooth optimization with constraints, and we convert it to the following unconstrained

equivalence problem:

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{A}} \quad & \frac{1}{2} \|\mathbf{Y} - \mathbf{EA}\|_F^2 + \lambda \phi_{\text{vol}}(\mathbf{E}) + \alpha \|\nabla \mathbf{A}^T\|_1 \\ & + \beta \|\mathbf{W} \odot \mathbf{A}\|_1 + l_{\mathbf{R}^+}(\mathbf{E}) + l_{\mathbf{R}^+}(\mathbf{A}) + l_{\{1\}}(\mathbf{A}) \end{aligned} \quad (9)$$

where  $l_{\mathbf{R}^+}(\cdot)$  and  $l_{\{1\}}(\cdot)$  represents the ANC and ASC, respectively, and can be computed as

$$l_{\mathbf{R}^+}(\mathbf{A}) = \begin{cases} 0 & \min(\mathbf{A}_i) \geq 0, \quad i = 1, 2, \dots, N \\ +\infty & \min(\mathbf{A}_i) < 0, \quad i = 1, 2, \dots, N \end{cases} \quad (10)$$

$$l_{\{1\}}(\mathbf{A}) = \begin{cases} 0 & \sum \mathbf{A}_i = 1, \quad i = 1, 2, \dots, N \\ +\infty & \sum \mathbf{A}_i \neq 1, \quad i = 1, 2, \dots, N. \end{cases} \quad (11)$$

### A. Variable Splitting and Augmented Lagrangian Approach

In order to effectively solve (9), we introduce the auxiliary variables  $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4$ , and convert (9) into the following constraint forms:

$$\begin{aligned} \min_{\mathbf{V}, \mathbf{E}, \mathbf{A}} \quad & \frac{1}{2} \|\mathbf{Y} - \mathbf{EA}\|_F^2 + \lambda \phi_{\text{vol}}(\mathbf{E}) + \alpha \|\mathbf{V}_1\|_1 \\ & + \beta \|\mathbf{W} \odot \mathbf{V}_2\|_1 + l_{\mathbf{R}^+}(\mathbf{E}) + l_{\mathbf{R}^+}(\mathbf{V}_3) \\ & + l_{\{1\}}(\mathbf{V}_4) \\ \text{s.t.} \quad & \mathbf{V}_1 = \nabla \mathbf{A}^T, \mathbf{V}_2 = \mathbf{A}, \mathbf{V}_3 = \mathbf{A}, \mathbf{V}_4 = \mathbf{A}. \end{aligned} \quad (12)$$

Equation (12) is a typical equality constraint optimization, and its augmented Lagrangian function can be formulated as follows:

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{V}, \mathbf{E}, \mathbf{A}) = \quad & \frac{1}{2} \|\mathbf{Y} - \mathbf{EA}\|_F^2 + \lambda \phi_{\text{vol}}(\mathbf{E}) + l_{\mathbf{R}^+}(\mathbf{E}) \\ & + \alpha \|\mathbf{V}_1\|_1 + \beta \|\mathbf{W} \odot \mathbf{V}_2\|_1 + l_{\mathbf{R}^+}(\mathbf{V}_3) \\ & + l_{\{1\}}(\mathbf{V}_4) + \frac{\mu}{2} \|\mathbf{V}_1 - \nabla \mathbf{A}^T - \mathbf{D}_1\|_F^2 \\ & + \frac{\mu}{2} \|\mathbf{V}_2 - \mathbf{A} - \mathbf{D}_2\|_F^2 \\ & + \frac{\mu}{2} \|\mathbf{V}_3 - \mathbf{A} - \mathbf{D}_3\|_F^2 \\ & + \frac{\mu}{2} \|\mathbf{V}_4 - \mathbf{A} - \mathbf{D}_4\|_F^2 \end{aligned} \quad (13)$$

where  $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$ , and  $\mathbf{D}_4$  are the scaled Lagrange multipliers, and  $\mu > 0$  is a penalty parameter. We choose to minimize the augmented Lagrangian function with respect to each block variable  $\mathbf{V}_1, \dots, \mathbf{V}_4, \mathbf{A}$ , and  $\mathbf{E}$  one at a time while fixing the other five blocks at their latest values, and then update the Lagrange multiplier. In specific, we optimize (13) by iteratively solving

the following unconstrained subproblems:

$$\left\{ \begin{array}{l} \mathbf{V}_1^{k+1} = \arg \min_{\mathbf{V}_1 \in \mathbb{R}^{N \times d}} \mathcal{L}_\mu \\ \quad (\mathbf{V}_1, \mathbf{V}_2^k, \mathbf{V}_3^k, \mathbf{V}_4^k, \mathbf{E}^k, \mathbf{A}^k), \quad (14a) \\ \mathbf{V}_2^{k+1} = \arg \min_{\mathbf{V}_2 \in \mathbb{R}^{d \times N}} \mathcal{L}_\mu \\ \quad (\mathbf{V}_1^{k+1}, \mathbf{V}_2, \mathbf{V}_3^k, \mathbf{V}_4^k, \mathbf{E}^k, \mathbf{A}^k) \quad (14b) \\ \mathbf{V}_3^{k+1} = \arg \min_{\mathbf{V}_3 \in \mathbb{R}^{d \times N}} \mathcal{L}_\mu \\ \quad (\mathbf{V}_1^{k+1}, \mathbf{V}_2^{k+1}, \mathbf{V}_3, \mathbf{V}_4^k, \mathbf{E}^k, \mathbf{A}^k) \quad (14c) \\ \mathbf{V}_4^{k+1} = \arg \min_{\mathbf{V}_4 \in \mathbb{R}^{d \times N}} \mathcal{L}_\mu \\ \quad (\mathbf{V}_1^{k+1}, \mathbf{V}_2^{k+1}, \mathbf{V}_3^{k+1}, \mathbf{V}_4, \mathbf{E}^k, \mathbf{A}^k) \quad (14d) \\ \mathbf{A}^{k+1} = \arg \min_{\mathbf{A} \in \mathbb{R}^{d \times N}} \mathcal{L}_\mu \\ \quad (\mathbf{V}_1^{k+1}, \mathbf{V}_2^{k+1}, \mathbf{V}_3^{k+1}, \mathbf{V}_4^{k+1}, \mathbf{A}, \mathbf{E}^k) \quad (14e) \\ \mathbf{E}^{k+1} = \arg \min_{\mathbf{E} \in \mathbb{R}^{L \times d}} \mathcal{L}_\mu \\ \quad (\mathbf{V}_1^{k+1}, \mathbf{V}_2^{k+1}, \mathbf{V}_3^{k+1}, \mathbf{V}_4^{k+1}, \mathbf{A}^{k+1}, \mathbf{E}) \quad (14f) \\ \mathbf{D}_1^{k+1} = \mathbf{D}_1^k + \nabla \mathbf{A}^{(k+1)T} - \mathbf{V}_1^{k+1} \quad (14g) \\ \mathbf{D}_i^{k+1} = \mathbf{D}_i^k + \mathbf{A}^{k+1} - \mathbf{V}_i^{k+1}, i = 2, \dots, 4. \quad (14h) \end{array} \right.$$

### B. Subproblems Solver

The  $\mathbf{V}_1$  and  $\mathbf{V}_2$  subproblem can be effectively solved by the well-known *soft-thresholding* approach [35], which is defined as

$$\text{softTh}(x, \tau) = \text{sign}(x) \max\{0, |x| - \tau\}, \text{ and } \tau > 0.$$

Then, we can get the following update rules:

$$\mathbf{V}_1^{k+1} \leftarrow \text{softTh} \left( \nabla \mathbf{A}^{(k)T} + \mathbf{D}_1^k, \frac{\alpha}{\mu} \right) \quad (15)$$

$$\mathbf{V}_2^{k+1} \leftarrow \text{softTh} \left( \mathbf{A}^k + \mathbf{D}_2^k, \frac{\beta}{\mu} \mathbf{W}^k \right) \quad (16)$$

where  $\mathbf{W}$  is the weighted matrix defined by (6).

In  $\mathbf{V}_3$  subproblem, the role of the  $l_{\mathbf{R}^+}$  term is to project the solution onto the nonnegative orthant and the value of  $\mathbf{V}_3$  is given by

$$\mathbf{V}_3^{k+1} \leftarrow \max(\mathbf{A}^k + \mathbf{D}_3^k, 0). \quad (17)$$

The  $\mathbf{V}_4$  subproblem, according to the definition of (11), it can be solved by

$$\mathbf{V}_4^{k+1} \leftarrow (\mathbf{A}^k + \mathbf{D}_4^k) + (\mathbf{1}_d \otimes \mathbf{M}) \quad (18)$$

where  $\mathbf{1}_d = \text{ones}(d, 1)$  and  $\mathbf{M} = \frac{1}{d}[\text{ones}(1, N) - \sum_{j=1}^N (\mathbf{A}^k + \mathbf{D}_4^k)]$  and the  $\otimes$  is denoted as the Kronecker product.

The  $\mathbf{A}$  subproblem is a least squares (LS) problem, then the solution can be obtained by setting the partial derivative  $\Psi(\mathbf{A})$  as zero, i.e.,

$$\begin{aligned} \Psi(\mathbf{A}) &= \frac{1}{2} \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \frac{\mu}{2} \|\mathbf{V}_1 - \nabla \mathbf{A}^T - \mathbf{D}_1\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_2 - \mathbf{A} - \mathbf{D}_2\|_F^2 + \frac{\mu}{2} \|\mathbf{V}_3 - \mathbf{A} - \mathbf{D}_3\|_F^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{V}_4 - \mathbf{A} - \mathbf{D}_4\|_F^2 \end{aligned}$$

$$\begin{aligned} \partial_{\mathbf{A}} \Psi(\mathbf{A}) &= \mathbf{E}^T \mathbf{E} \mathbf{A} + \mu \mathbf{A} \nabla^T \nabla + 3\mu \mathbf{A} - \text{RHS} = 0 \\ &\Rightarrow \mathbf{E}^T \mathbf{E} \mathbf{A} + \mu \mathbf{A} \nabla^T \nabla + 3\mu \mathbf{A} = \text{RHS} \end{aligned} \quad (19)$$

where

$$\begin{aligned} \text{RHS} &= \mathbf{E}^T \mathbf{Y} + \mu [(\mathbf{V}_1 - \mathbf{D}_1)^T \nabla + (\mathbf{V}_2 - \mathbf{D}_2) \\ &\quad + (\mathbf{V}_3 - \mathbf{D}_3) + (\mathbf{V}_4 - \mathbf{D}_4)]. \end{aligned}$$

This equation can be rewritten as

$$\begin{aligned} [\text{vec}(\mathbf{I}_N \otimes \mathbf{E}^T \mathbf{E}) + \mu * \text{vec}(\nabla^T \nabla \otimes \mathbf{I}_d) + 3\mu] \text{vec}(\mathbf{A}) \\ = \text{vec}^1(\text{RHS}) \end{aligned} \quad (20)$$

and can be efficiently solved by the PCG [36] or LSQR [37].

The  $\mathbf{E}$  subproblem is defined as

$$\begin{aligned} J(\mathbf{E}) &= \min_{\mathbf{E}} \mathcal{L}_\mu(\mathbf{V}_i^{k+1}, \mathbf{A}^{k+1}, \mathbf{E}), i = 1, 2, \dots, 4 \\ &= \frac{1}{2} \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \lambda \phi_{\text{vol}}(\mathbf{E}) + l_{\mathbf{R}^+}(\mathbf{E}) \end{aligned}$$

where  $\phi_{\text{vol}}(\mathbf{E})$  is given by (3). Obviously,  $J(\mathbf{E})$  is a smooth convex optimization with nonnegative constraints. Intuitively, the projection gradient (PG) [36] learning scheme can be used to solve it. Thus, the update rule is expressed as

$$\mathbf{E}^{k+1} \leftarrow \mathbf{P}(\mathbf{E}^k - \rho^k \nabla J(\mathbf{E}^k)) \quad (21)$$

where  $\rho^k$  is stepsize and the operator  $\mathbf{P}(\cdot)$  projects all the negative entries to zero. To avoid inefficient line searches, we use the Barzilai–Borwein (BB) gradient method [38] based on two-point steps to select  $\rho$  as shown below

$$\rho^k = \frac{\langle \Delta_E, \Delta_E \rangle}{\langle \Delta_E, \Delta_g \rangle} \text{ or } \rho^k = \frac{\langle \Delta_E, \Delta_g \rangle}{\langle \Delta_g, \Delta_g \rangle} \quad (22)$$

where  $\Delta_E = \mathbf{E}^k - \mathbf{E}^{k-1}$ ,  $\Delta_g = \nabla J(\mathbf{E}^k) - \nabla J(\mathbf{E}^{k-1})$ ,  $\langle \cdot, \cdot \rangle$  is inner product.

### C. Overall Algorithm for MPEC-NMF

By incorporating the several steps described above, we can now present the entire procedure of solving the proposed model, as summarized in Algorithm 1.

### D. Implementation Details

Due to the global nonconvexity of the proposed MPEC-NMF model, some implementation details of the method, such as initialization and parameter values, need to be specified.

The first issue concerns the initialization of the  $\mathbf{E}$ ,  $\mathbf{A}$ ,  $\mathbf{W}$  and  $\mathbf{V}_1, \dots, \mathbf{V}_4$ , since an inappropriate initialization may cause the algorithm to be stuck in local minima. For the matrix  $\mathbf{E}$ , there are usually two initialization strategies: random initialization or selection from the original data points. Random initialization means randomly assigning a value between 0 and 1 as an element of the  $\mathbf{E}$ . The latter typically uses an unsupervised endmember extraction algorithm result as an input to  $\mathbf{E}$ , such as vertex

<sup>1</sup>The operation  $\text{vec}(\cdot)$  represents vectorization operation on matrix with columns stack.

**Algorithm 1:** MPEC-NMF Algorithm.

---

**Input:**  $\mathbf{Y} \in \mathfrak{R}^{L \times N}$ ,  $d, \lambda, \alpha, \beta, \mu$  and MaxIter.  
**Output:**  $\mathbf{E} \in \mathfrak{R}^{L \times d}$ ,  $\mathbf{A} \in \mathfrak{R}^{d \times N}$

**Initialization:** Set  $k = 0$ , choose  $\mu > 0$ ,  $\mathbf{E}^0 \in \mathfrak{R}^{L \times d}$ ,  
 $\mathbf{A}^0 \in \mathfrak{R}^{d \times N}$ ,  $\mathbf{W}_{i,j}^0 = 1/(|\mathbf{A}_{i,j}^0|^{1/2} + \epsilon)$ ,  
 $\mathbf{V}_1^0 = \nabla \mathbf{A}^{0T}$ ,  $\mathbf{V}_2^0 = \mathbf{A}^0$ ,  $\mathbf{V}_3^0 = \mathbf{A}^0$ ,  
 $\mathbf{V}_4^0 = \mathbf{A}^0$ ,  $\mathbf{D}_1^0 = \text{zeros}(\text{size}(\mathbf{A}^T))$ ,  $\mathbf{D}_2^0 = \mathbf{D}_3^0 =$   
 $\mathbf{D}_4^0 = \text{zeros}(\text{size}(\mathbf{A}))$ .

- 1: **repeat:**
- 2:  $\mathbf{V}_1^{k+1} \leftarrow \text{softTh}(\nabla \mathbf{A}^{(k)T} + \mathbf{D}_1^k, \frac{\alpha}{\mu})$ .
- 3:  $\mathbf{V}_2^{k+1} \leftarrow \text{softTh}(\mathbf{A}^k + \mathbf{D}_2^k, \frac{\beta}{\mu} \mathbf{W}^k)$ .
- 4:  $\mathbf{V}_3^{k+1} \leftarrow \max(\mathbf{A}^k + \mathbf{D}_3^k, 0)$ .
- 5:  $\mathbf{V}_4^{k+1} \leftarrow (\mathbf{A}^k + \mathbf{D}_4^k) + (\mathbf{1}_d \otimes \mathbf{M})$ , where  
 $\mathbf{1}_d = \text{ones}(d, 1)$  and  
 $\mathbf{M} = 1/d[\text{ones}(1, N) - \sum_{j=1}^N (\mathbf{A}^k + \mathbf{D}_4^k)]$ .
- 6:  $\mathbf{A}^{k+1} \leftarrow$  Solved by PCG from (20).
- 7:  $\mathbf{E}^{k+1} \leftarrow \mathbf{P}(\mathbf{E}^k - \rho^k \nabla J(\mathbf{E}^k))$ , where  $J(\mathbf{E}) =$   
 $1/2 \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \lambda \phi_{\text{vol}}(\mathbf{E}) + l_{\mathbf{R}^+}(\mathbf{E})$ .
- 8: **Update multipliers:**
- 9:  $\mathbf{D}_1^{k+1} \leftarrow \mathbf{D}_1^k + \nabla \mathbf{A}^{(k+1)T} - \mathbf{V}_1^{k+1}$ ,
- 10:  $\mathbf{D}_i^{k+1} \leftarrow \mathbf{D}_i^k + \mathbf{A}^{k+1} - \mathbf{V}_i^{k+1}, i = 2, \dots, 4$ .
- 11: **Update weight matrix:**
- 12:  $\mathbf{W}_{i,j}^{k+1} \leftarrow 1/(\sqrt{|\mathbf{A}_{i,j}^k|} + \epsilon)$ .
- 13: **Update iteration:**
- 14:  $k \leftarrow k+1$ .
- 15: **until** some stopping criterion is satisfied.
- 16: **return** results

---

component analysis (VCA) or spectral information divergence (SID). With this initialization, we then use the fully constrained least squares (FCLS) [39] to generate the abundance  $\mathbf{A}$ . The weighted matrix is initialized as  $\mathbf{W}_{i,j}^0 = 1/(|\mathbf{A}_{i,j}^0|^{1/2} + \epsilon)$  for the first iteration. As for the auxiliary variables, we have experimentally found that simply taking  $\mathbf{V}_1^0 = \nabla \mathbf{A}^{0T}$ ,  $\mathbf{V}_2^0 = \mathbf{A}^0$ ,  $\mathbf{V}_3^0 = \mathbf{A}^0$ ,  $\mathbf{V}_4^0 = \mathbf{A}^0$  is an effective choice.

The second issue is about the model parameters:  $\lambda$ ,  $\alpha$ , and  $\beta$ . Three parameters control the tradeoff between the reconstruction error term and the three *prior*-learners.  $\lambda$  controls the strength of the volume constraint. Increasing  $\lambda$  can make the data cloud in the simplex more compact and weaken noise interference. Therefore,  $\lambda$  value should increase or decrease as the magnitude of the noise level changes. The parameter  $\alpha$  controls the TV regularization, which forces abundance piecewise smooth. It is empirical and the selection is dependent on the specific image. The value of  $\beta$  relies on the sparsity levels of the material abundance. Since the abundance cannot be known in advance, we use a possible estimator based on the sparse coding [40] as a reference, given by

$$\beta = \frac{1}{\sqrt{L}} \sum_l \frac{\sqrt{N} - \|\mathbf{y}_l\|_1 / \|\mathbf{y}_l\|_2}{\sqrt{N} - 1} \quad (23)$$

where  $\mathbf{y}_l$  represents the given hyperspectral data at band  $l$ . It is important to note that adaptive selection of optimal values for

all these parameters is still an open question. In the next section, we analyze the effects of these parameters through experiments.

The  $\mu$  is a penalty parameter of the augmented Lagrangian, and its setting has a strong influence on the convergence speed of the algorithm. Here, we use an adaptive strategy [41]. Specifically,  $\mu$  is updated by keeping the ratio between the ADMM primal residual norm and dual residual norm within a given positive interval, since they both converge to zero. We use the following primal residual norm and dual residual norm definition to measure the degree of iteration satisfaction of Algorithm 1:

$$r^k = \left\| [\mathbf{V}_1^k; \mathbf{V}_2^k; \mathbf{V}_3^k; \mathbf{V}_4^k] \right\|_F^2 - \left\| [(\nabla \mathbf{A}^T)^k; \mathbf{A}^k; \mathbf{A}^k; \mathbf{A}^k] \right\|_F^2 \quad (24)$$

$$d^k = \left\| [\mathbf{V}_1^k; \mathbf{V}_2^k; \mathbf{V}_3^k; \mathbf{V}_4^k] \right\|_F^2 - \left\| [\mathbf{V}_1^{k-1}; \mathbf{V}_2^{k-1}; \mathbf{V}_3^{k-1}; \mathbf{V}_4^{k-1}] \right\|_F^2 \quad (25)$$

In experiment, the primal residual  $r^k$  and the dual residual  $d^k$  were set to  $1e-3$ . In addition to the termination criteria above, we also set a maximum number of iterations. The optimization procedure stops as long as either of the two criteria is met. We have empirically found that the combination of the above two strategies performs very well in our case.

The last issue concerns the estimation of the endmember number. Even though this is important for unmixing task, it is another independent topic that is not covered in this work. In fact, two well-known methods such as VD [42] and HySime [43] could be adopted to estimate the number of endmembers. In the synthetic data experiment, we manually set the number values, and in the real data, we refer to the results of previous hyperspectral community studies.

### E. Complexity Analysis

Here, we briefly analyze the computational complexity of the proposed MPEC-NMF algorithm. The most costly steps of the proposed method are to solve  $\mathbf{A}$  and  $\mathbf{E}$  subproblem (i.e., steps 6 and 7 in *Algorithm 1*), while the others computational complexity is  $O(N)$  [44].

The  $\mathbf{A}$  subproblem is a LS problem, and its computational complexity mainly consists of calculating  $\partial_{\mathbf{A}} \Psi(\mathbf{A})$  and executing the PCG algorithm. The  $\partial_{\mathbf{A}} \Psi(\mathbf{A})$  part has a complexity of  $O(dN^2)$  because it contains TV operations. The PCG is an improved variant of the CG method [36]. The time complexity of the CG approach is  $O(mk^{1/2})$  [45], where  $m$  and  $k$  are the number of nonzero elements and the condition number of the LS coefficient matrix, respectively. In our algorithm, since the partial derivative equation for  $\mathbf{A}$  contains a TV operation, the coefficient matrix size of  $\mathbf{A}$  is  $dN \times dN$ , which is a large sparse matrix, thus,  $m \ll d^2 N^2$  and is a small constant. Taking all operations into account, the computational complexity order for solving the  $\mathbf{A}$  subproblem is  $O(dN^2 + mk^{1/2})$  per iteration.

The  $\mathbf{E}$  subproblem is a smooth convex optimization with non-negative constraints. In step 7, the projected Barzilai–Borwein (PBB) gradient method [46], [47] is adopted. The Barzilai–Borwein (BB) gradient method [38] has a R-linear convergence

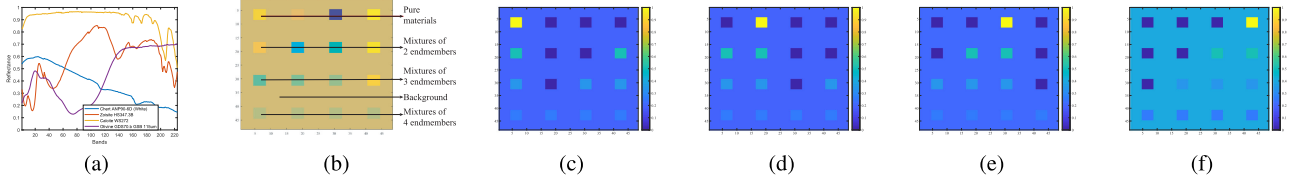


Fig. 4. Synthetic hyperspectral data. (a) Reference endmembers. (b) Simulated image. (c)–(f) Ground-truth fractional abundances of corresponding endmembers.

rate [48], [49] and is competitive with the CG method. In the PBB method, the most costly steps include calculating  $\nabla J(\mathbf{E}^k)$ , simplex volume, Barzilai–Borwein stepsize  $\rho$ , and nonnegative elements projection operations. Among them, except that the computational complexity of  $\nabla J(\mathbf{E}^k)$  is  $O(dLN)$ , the remaining complexity is  $O(dL)$ . Taking all the operations into consideration, the computational complexity order of the  $\mathbf{E}$  subproblem is  $O(dLN + 3dL)$  per iteration.

Based on the above analysis, the computational complexity of the proposed MPEC-NMF algorithm is  $O(dN^2 + mk^{1/2})$ , and the most time-consuming operation is to solve  $\mathbf{A}$  subproblem.

## V. EXPERIMENTAL RESULTS

In this section, the experiments on synthetic and real dataset are conducted to verify the effectiveness of the proposed method. The state-of-the-arts method including VCA-FCLS [6],<sup>2</sup> [39], MVC-NMF [12],<sup>3</sup>  $L_{1/2}$ -NMF [13], GLNMF [14], TV-RSNMF [15],<sup>4</sup> MVNTF-TV [25], and DLNMF-TV [10] are utilized as comparison methods. All the experiments are carried out on a Windows 7 system with 3.4-GHz Intel Core i7 CPU and 16-GB RAM using MATLAB2016b.

### A. Performance Metrics

The results were evaluated using the *Spectral Angle Distance* (SAD) and *Root Mean Square Error* (RMSE). They are defined as follows:

$$\text{SAD}_k = \arccos \left( \frac{\mathbf{e}_k^T \hat{\mathbf{e}}_k}{\|\mathbf{e}_k\| \cdot \|\hat{\mathbf{e}}_k\|} \right) \quad (26)$$

$$\text{RMSE}_k = \left( \frac{1}{N} \|\mathbf{a}_k - \hat{\mathbf{a}}_k\|_2^2 \right)^{1/2} \quad (27)$$

where  $\mathbf{e}_k$  and  $\mathbf{a}_k$  denote the ground-truth of endmember signature and abundance vector, respectively,  $\hat{\mathbf{e}}_k$  and  $\hat{\mathbf{a}}_k$  denote the corresponding estimated results and  $N$  represents the number of pixels in HSI.

### B. Experiment on Synthetic Dataset

In this section, we randomly selected four spectra from the U.S. Geological Survey (USGS) digital spectral library,<sup>5</sup> as shown in Fig. 4(a), for the following experiments. The synthetic data are shown in Fig. 4(b), which was generated using the

method provided in [50]. This simulated image has  $48 \times 48$  pixels with 188 bands were generated using four spectral signatures according to LMM, and applied ASC to the pixels. There are not only pure pixel regions in the image, but also mixed regions constructed using mixtures ranging two and four endmembers. Background pixels are also constructed by mixtures of the same four endmembers, and their corresponding fractional abundances are fixed at 0.1870, 0.2023, 0.2036, and 0.4071, respectively. In addition, the true fractional abundances of the corresponding endmembers have been shown in Fig. 4(c)–(f).

In order to evaluate the sensitivity of the method to noise, we add zero-mean Gaussian noise to the synthetic data to simulate possible errors and noise interference. Here, the SNR is defined as follows:

$$\text{SNR}_{\text{db}} = 10 \log_{10} \frac{E[\mathbf{y}^T \mathbf{y}]}{E[\mathbf{g}^T \mathbf{g}]} \quad (28)$$

where  $\mathbf{y}$  and  $\mathbf{g}$  represent the observation and noise of pixel, respectively, and  $E[\cdot]$  denotes the expectation operator.

In order to guarantee a fair comparison of all methods, unless otherwise stated, the initial endmember matrix of all NMF-based methods is initialized with the SID, and the initial abundance is randomly generated and normalized by the *randn* function using the following code: “ $A = \text{abs}(\text{randn}(d,N)); A = A ./ \text{repmat}(\text{sum}(A,1),d,1);$ ”

All experimental results in this article are average values obtained from 20 randomized tests. In addition, all parameters of the comparison algorithm are consistent with the original references.

*Experiment 1 (Parameter Analysis):* In this experiment, three critical parameters  $\lambda$ ,  $\alpha$ , and  $\beta$  in model (8) are considered when SNR = 20 dB. In order to eliminate the influence of random initialization, the initial  $\mathbf{E}$  and  $\mathbf{A}$  values were kept fixed throughout the experiment, and the initial augmented Lagrangian penalty parameter  $\mu = 1$ , the maximum number of iterations were 350, and the error tolerance was  $1e-3$ .

First, we analyze the parameter  $\lambda$  while fixing  $\alpha$  and  $\beta$  values. As shown in Fig. 5, when  $\lambda$  changes from 0.02 to 0.04, the SAD and RMSE values are relatively stable in this interval. When  $\lambda$  is below 0.02, both SAD and RMSE values rise clearly. Therefore, by weighing the best SAD and RMSE results, we choose the best parameter  $\lambda = 0.025$  for subsequent experiments.

Next, we evaluated the effects of different  $\alpha$  and  $\beta$  values on performance. In this experiment, the values of parameter  $\alpha$  are set to a finite set  $\{1e-3, 5e-3, 1e-2, 1.5e-2, 2e-2, 2.5e-2, 3e-2, 4e-2\}$  according to the comprehensive analysis of the existing research results. For parameter  $\beta$ , we set it as  $\{1e-4, 1e-3, 2e-3, 3e-3, 4e-3, 5e-3, 6e-3, 7e-3\}$ , which is the

<sup>2</sup><http://www.lx.it.pt/~bioucas/code.htm>

<sup>3</sup><https://github.com/aicp/MVCNMF>

<sup>4</sup><https://sites.google.com/site/rshewei/home>

<sup>5</sup><http://speclab.cr.usgs.gov/spectral.lib06>



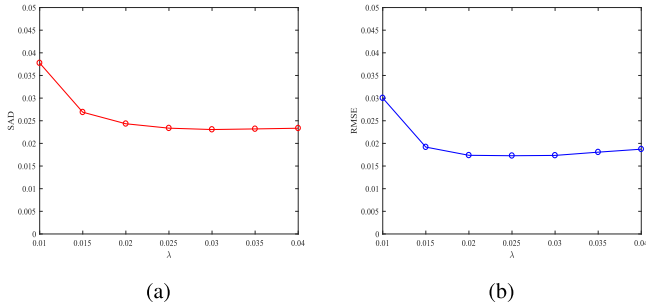


Fig. 5. Performance analysis of the proposed method with respect to parameters  $\lambda$  in terms of (a) SAD and (b) RMSE.

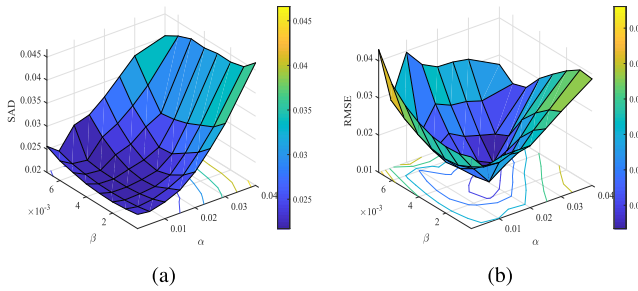


Fig. 6. Performance analysis of the proposed method with respect to parameters  $\alpha$  and  $\beta$  in terms of (a) SAD and (b) RMSE.

result of fine-tuning under the guidance of sparse estimator (23). During the experiment,  $\lambda$  value was fixed, and all possible combinations of values of  $\alpha$  and  $\beta$  were considered.

Fig. 6(a) shows the SAD obtained by the proposed method when  $\alpha$  and  $\beta$  parameters are changed at  $\lambda = 0.025$ , and Fig. 6(b) shows the RMSE results. It can be clearly observed that when the  $\alpha$  value exceeds 0.02, the SAD and RMSE values have a clear increasing trend. Similarly, when  $\alpha$  value is below 0.01, the RMSE value also rises rapidly. In contrast, the two metrics achieved the lowest SAD and RMSE values at  $\alpha = 0.015$ . At the same time, when  $\beta$  exceeds 0.002 and is lower than 0.004, the SAD and RMSE values are stable and robust. For other values, whether larger or smaller, the values of both metrics show a rising tendency. Therefore, the optimal value of  $\beta$  should be in  $[0.002, 0.004]$ . Based on the above analysis, we set  $\beta = 0.003$  and  $\alpha = 0.015$  in the rest of the experiments.

*Experiment 2 (Performance Comparisons):* The main goal of this experiment is to evaluate the robustness of the proposed method at different noise levels, i.e., SNR = 10, 20, 30, and 40 dB. It should be noted that all the methods use the same initialization conditions. Specifically, the initial endmember matrix of all NMF-based methods is initialized with the SID, and the abundance matrix is initialized in a random manner to better assess the ability of each algorithm to estimate abundance. The experimental results are shown in Fig. 7.

Obviously, with the increase of SNR, the SAD and RMSE values obtained by all methods show a rapid decline trend, which indicates that the performance of all algorithms is increasing, and it also implies the importance of noise reduction in the unmixing scenarios. It can be observed from Fig. 7(a) and (b)

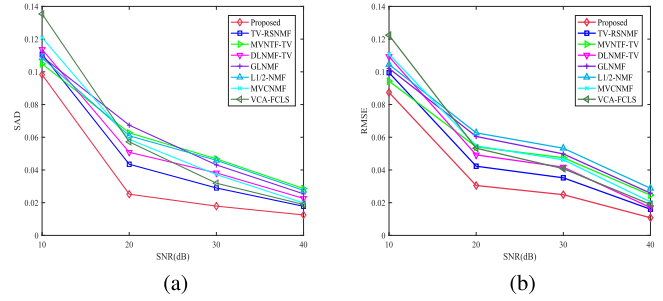


Fig. 7. Comparison of the algorithms at different noise levels in terms of (a) SAD and (b) RMSE.

TABLE I  
AVERAGE PROCESSING TIME (IN SECONDS) OF EACH COMPETING METHOD ON SIMULATED DATASETS

Methods	Proposed	TV-RSNMF	MVNTE-TV	DLNMF-TV	GLNMF	L1/2-NMF	MVCNMF	VCA-FCLS
Time[s]	27.36	208.4	263.72	229.15	335.5	349.34	42.64	0.89

that, due to the existence of pure pixels in the simulated data, the VCA method based on subspace projection can also obtain better unmixing results than most algorithms in clean images, but its performance shows rapid degradation as the noise intensity increases, and it gets the worst results when SNR = 10 dB. In contrast, under the random initialization of the abundance matrix, all NMF-type methods show similar estimation capabilities with changes in noise levels, and it can be seen from two measurement curves of SAD and RMSE that they have stronger stability than geometric-class methods. However, these approaches only consider geometric and statistical *priors* in isolation, and only impose constraints on endmember or abundance in a single aspect, ignoring the effects of multiple *priors* joint and cooperation constraints of two matrix factors.

Interestingly, as shown in Fig. 7(a) and (b), regardless of the noise intensity, our proposed approach achieves the lowest SAD and RMSE metrics than other methods, showing better stability and robustness. Therefore, it further proves our previous hypothesis, namely, by fusing geometric and statistical *prior* information, ensemble multiple *priors* regularizer imposes constraints on the endmember and the abundance matrix, which can shrink the solution space more effectively.

For illustrative purposes, Fig. 8 presents the estimated abundance maps for randomly selected first endmember. It can be clearly seen that no matter the noise level, our proposed method can always obtain smoother and clearer abundance due to considering multiple *prior* constraints, and whether the background or foreground is more consistent with the real abundance map, showing its superiority and advancedness. Specifically, at low SNR, our method is significantly better than other competing methods, as shown in the first two lines of Fig. 8. In the case of high SNR, all methods have similar behavior and visual effects, but our method yields smoother textures and edges that closely match the true abundance.

In addition, we also give the average time cost of each algorithm to evaluate its time efficiency. As shown in Table I, it is not surprising that the VCA method has the fastest running speed because it extracts endmembers based on spatial projection without



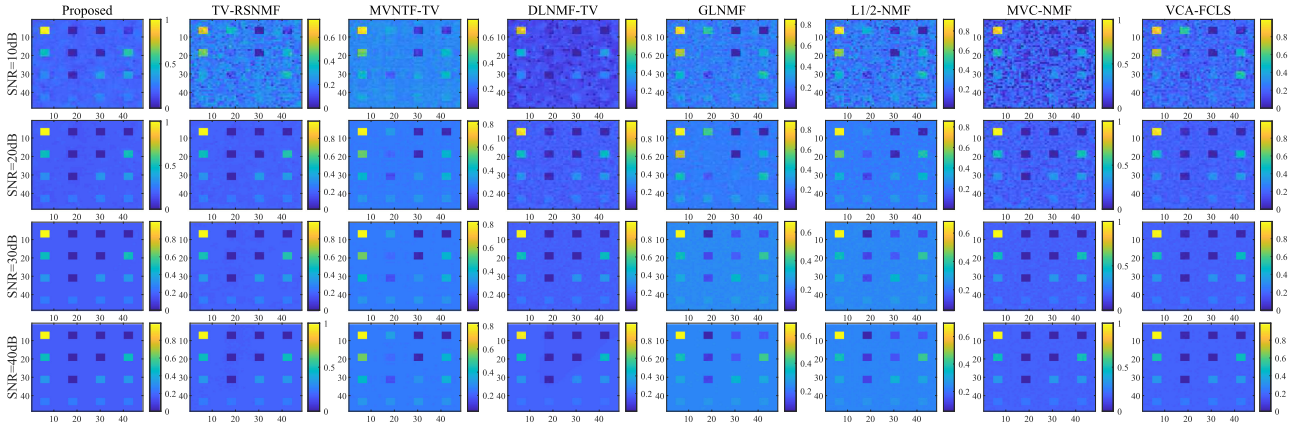


Fig. 8. Abundance maps of the first endmember obtained by different methods in the synthetic data experiment.

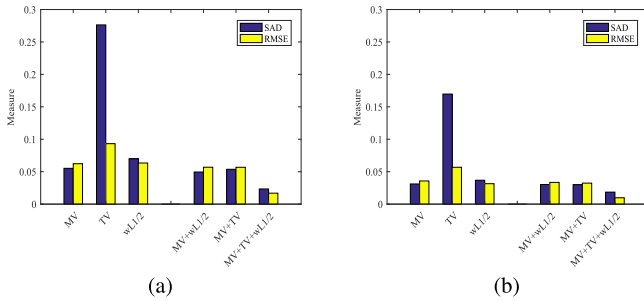


Fig. 9. Effects of different *priors* on performance in the proposed model under different initialization conditions. (a) Random. (b) LS.

the need to perform iterative optimization algorithms. Comparing with other NMF-class methods, the proposed approach has less time consumption. Because our proposed method adopts multiple *priors* ensemble constraints, and applies both geometric and statistical *priors* to the endmember and abundance matrices, thereby effectively compressing the objective solution space and improving the algorithm time efficiency. Moreover, the algorithm proposed in this article is based on an efficient ADMM framework combined with the BB gradient strategy, which also further accelerates the convergence rate.

*Experiment 3 (Contributions of different priors):* In this experiment, we explore the effects of different *prior* regularizers on the results of the algorithm. In order to further evaluate the characteristics of the proposed method, the abundance matrix  $\mathbf{A}$  is initialized in two ways, namely random method and LS method. In order to examine the contribution of each *prior* and its combination on the unmixing model, we invalidate the other *priors* when evaluating one of them.

Obviously, in Fig. 9(a), the first three group histograms showed that any single matrix constraint does not achieve the desired result, but the performance of the MV constraint is clearly better than the other two. By contrast, the SAD and RMSE values of the latter three groups are obviously lower than any of the previous groups. Moreover, the joint MV, weighted  $L_{1/2}$ , and TV regularization scheme (MV+w $L_{1/2}$ +TV) proposed in this article achieved lowest values. In order to comprehensively compare the contributions and influences of the *priors*, we present the test

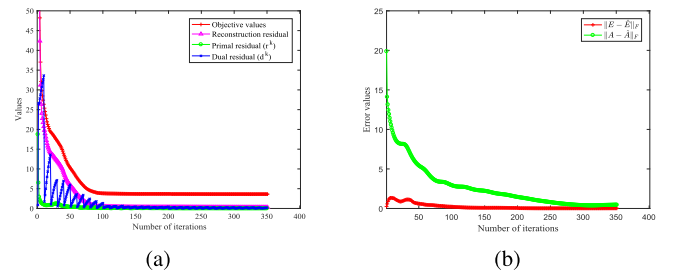


Fig. 10. Convergence curves. (a) Objective values, reconstruction, and primal-dual residuals. (b) Estimate the residuals of  $\hat{\mathbf{E}}$  and  $\hat{\mathbf{A}}$ .

results using the LS to initialize  $\mathbf{A}$ , as shown in Fig. 9(b). As can be seen, good initial points can improve the accuracy of all test groups. However, the results of the MV+w $L_{1/2}$ +TV scheme are still better than the others. From the comparative analysis of Fig. 9(a) and (b), it can be seen that the MV *prior* in the model has a larger influence and contribution. The weighted  $L_{1/2}$  sparsity (w $L_{1/2}$ ) contribution to the model is second only to the MV. The TV regularization mainly contributes to the abundance extraction, which is consistent with its theory. The values of the two metrics obtained by the three *priors* ensemble constraints (MV+w $L_{1/2}$ +TV) are superior to any other groups, which exert the ensemble effectiveness of multiple-*priors* learning, and has strong robustness.

*Experiment 4 (Convergence analysis):* Our proposed *algorithm 1* use (24) and (25) as the stopping criterion, and its convergence can be analyzed and proved in a similar way to [41] and [34]. Here, we did it in an alternative way by plotting the convergence curve of the proposed algorithm. In Fig. 10(a), the objective function values, reconstruction residual, primal residual norm, and dual residual norm values were drawn. It can be seen that the objective value and reconstruction residual drop rapidly after the first few iterations and converges to the stationary value after about 100 iterations. The primal residual and the dual residual also converge to zero after about 150 iterations, which quite comply with our expected effect.

Meanwhile, we also give the iterative residual curves of the estimated matrices  $\hat{\mathbf{E}}$  and  $\hat{\mathbf{A}}$ , as shown in Fig. 10(b). It can be

TABLE II  
SAD VALUES OF THE DIFFERENT ALGORITHMS WITH THE HYDICE URBAN DATASET

Algorithms	Proposed	TV-RSNMF	MVNTF-TV	DLNMF-TV	GLNMF	$L_{1/2}$ -NMF	MVC-NMF	VCA-FCLS
Asphalt-road	<b>0.1096</b>	0.1208	0.1462	0.1532	0.1597	0.1429	0.1194	0.2418
Grass	<b>0.0890</b>	0.0903	0.1009	0.0973	0.0991	0.0985	0.1410	0.2104
Tree	0.1277	0.1350	<b>0.1116</b>	0.1369	0.1398	0.1342	0.2210	0.4011
Roof#1	<b>0.1007</b>	0.1013	0.1267	0.1010	0.1356	0.1373	0.2121	0.2606
Roof#2	<b>0.0798</b>	0.0817	0.0996	0.0922	0.0801	0.1017	0.1176	0.5150
Concrete-road	0.0730	<b>0.0725</b>	0.0894	0.0728	0.0960	0.0818	0.2250	0.2495
Mean	<b>0.0966</b>	0.1003	0.1124	0.1089	0.1184	0.1161	0.1727	0.3131

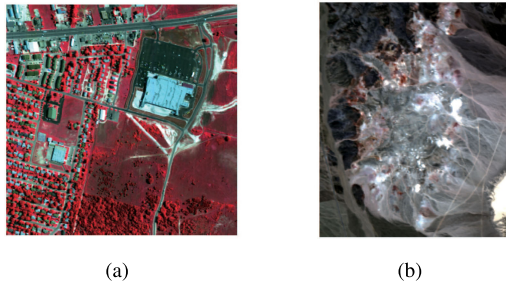


Fig. 11. Two real hyperspectral datasets. (a) HYDICE Urban scene (R: 69, G: 50, B: 36). (b) AVIRIS Cuprite scene (R: 28, G: 19, B: 11).

seen from the figure that after 300 iterations, the residuals of the two matrices tend to be stable and close to zero.

### C. Experiment on Real Dataset

Here, two real world HSI datasets were used to further evaluate the performance of the proposed method: the Hyperspectral Digital Image Acquisition Experiment (HYDICE) Urban Dataset<sup>6</sup> and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite dataset,<sup>7</sup> as shown in Fig. 11. The first scene was collected in October 1995 in an urban area at Copperas Cove, TX, USA. The second scene was collected in June 1997 in Las Vegas, NV, USA. In the experiment, we used a SID-based method to select pixels as the initial endmember matrix, all NMF-based methods have the same initialization conditions, and the experiment was repeated ten times to ensure a reliable comparison.

1) *HYDICE Urban Dataset*: This image has  $307 \times 307$  pixels and each pixel contains 210 bands, which cover the wavelength range of 0.4–2.4  $\mu\text{m}$ . By removing the noise and water vapor absorption bands (including bands 1–4, 76, 87, 101–111, 136–153, and 198–210), only 162 bands remain. According to previous research works [15], this scene consists of six materials, namely asphalt-road, roof#1, roof#2, grass, tree, and concrete-road.

Table II shows the mean SAD obtained by different methods. From this table, we can see that most of the material signatures extracted by proposed method have the best accuracy and the lowest SAD mean values. Another interesting finding is that the proposed method has better performance than any other methods, which imposed geometric or statistical *prior* constraints on a single matrix. Fig. 12 shows the estimated endmember

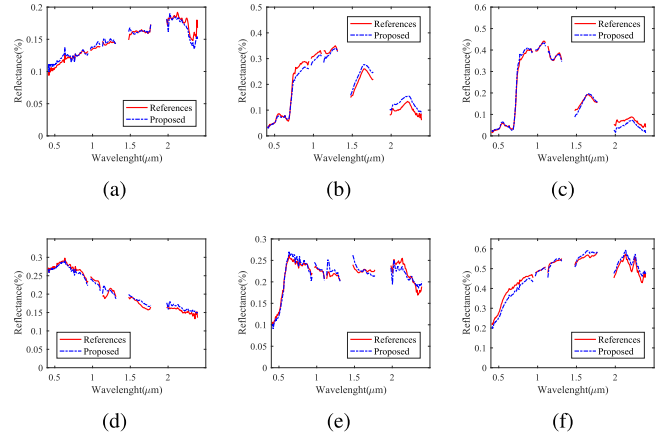


Fig. 12. Comparison of the reference spectra with the endmember signatures extracted by the proposed method on the Urban dataset. (a) Asphalt-road. (b) Grass. (c) Tree. (d) Roof#1. (e) Roof#2. (f) Concrete-road.

signatures and their references. It can be seen that the output of the proposed method is closely matched with the references signatures. Fig. 13(a)–(f) shows the abundance ground-truth of six endmembers compared to the estimated abundance maps in Fig. 13(g)–(l), which are grayscale image, where a dark pixel denotes a low abundance of the corresponding endmember. It can be seen that these results are quite reasonable comparing with references.

2) *AVIRIS Cuprite Dataset*: The Cuprite dataset contains 224 bands, which cover the wavelength range of 0.4–2.5  $\mu\text{m}$  with subimage size  $250 \times 191$ . The noisy bands (1 and 2 and 221–224) and water absorption bands (104–113 and 148–167) were removed, and a total of 188 bands remained. The reference endmember signatures in the experiment were selected from the USGS digital spectral library, which was also utilized in [15] and [13]. According to the existing analysis in [15] and [26], there are mainly 12 types minerals: “#1 Alunite GDS82 Na<sub>8</sub>,” “#2 Andradite WS487,” “#3 Buddingtonite GDS85 D-206,” “#4 Chalcedony CU91-6 A,” “#5 Dumortierite HS190.3B,” “#6 Kaolin/Smect H89-FR-5 30 K,” “#7 Kaolin/Smect KLF508 85%K,” “#8 Montmorillonite + Illi CM37,” “#9 Muscovite GDS108,” “#10 Nontronite NG-1.a,” “#11 Pyrope WS474,” and “#12 Sphene HS189.3B.”

Fig. 14 illustrates the grayscale abundance maps obtained by our proposed method, where the brighter pixels represent the higher abundance of the corresponding endmembers. Meanwhile, Fig. 15 presents the endmember signatures extracted by proposed method and the reference signatures obtained from

<sup>6</sup><http://www.agc.army.mil/>

<sup>7</sup>[http://aviris.jpl.nasa.gov/data/free\\_data.html](http://aviris.jpl.nasa.gov/data/free_data.html)

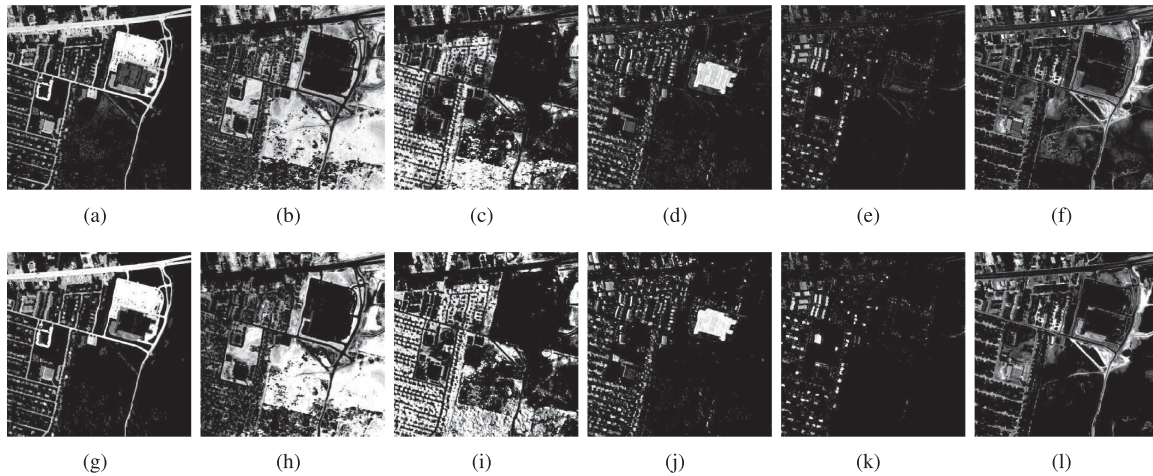


Fig. 13. Abundance maps of the different endmembers obtained using MPEC-NMF on the HYDICE Urban dataset. (a)–(f) Ground-truth. (g)–(l) Estimated. From left to right are: Asphalt-road, Grass, Tree, Roof#1, Roof#2, Concrete-road, respectively.

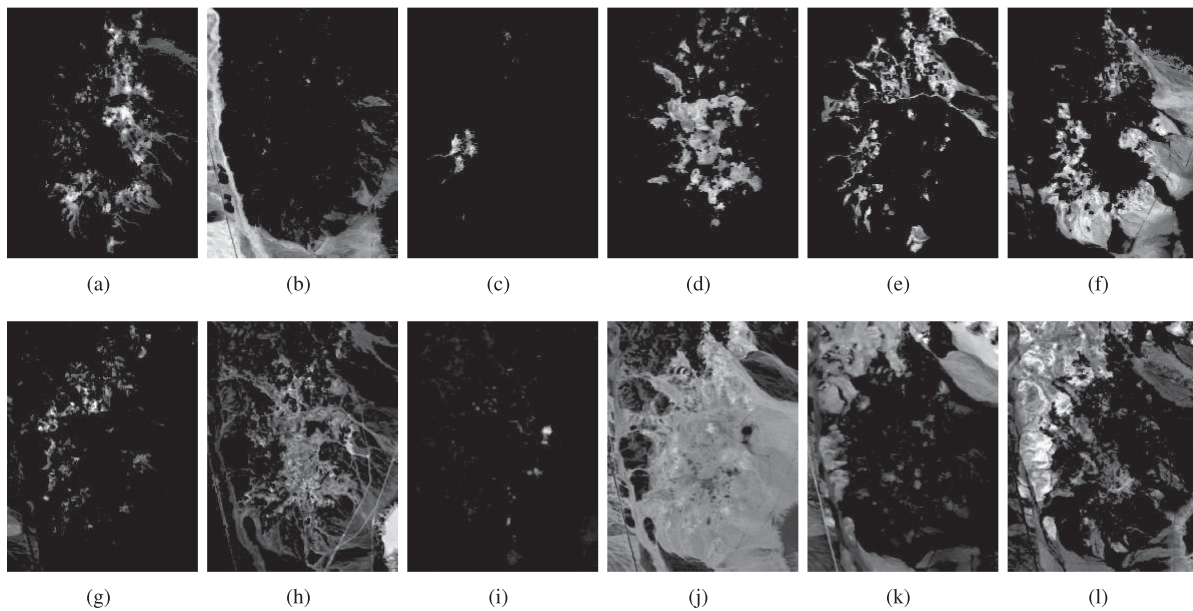


Fig. 14. Abundance maps of the different endmembers obtained using MPEC-NMF on the AVIRIS Cuprite dataset. (a) Alunite GDS82 Na82. (b) Andradite WS487. (c) Buddingtonite GDS85 D-206. (d) Chalcedony CU91-6 A. (e) Dumortierite HS190.3B. (f) Kaolin/Smect H89-FR-5 30 K. (g) Kaolin/Smect KLF508 85%K. (h) Montmorillonite + Illi CM37. (i) Muscovite GDS108. (j) Nontronite NG-1.a. (k) Pyrope WS474. (l) Sphene HS189.3B.

TABLE III  
SAD VALUES OF THE DIFFERENT ALGORITHMS WITH THE AVIRIS CUPRITE DATASET

Algorithms	Proposed	TV-RSNMF	MVNTF-TV	DLNMF-TV	GLNMF	$L_{1/2}$ -NMF	MVC-NMF	VCA-FCLS
Alunite GDS82 Na82	0.1241	0.1293	0.1239	0.1252	0.1246	0.1273	0.1730	<b>0.1224</b>
Andradite WS487	0.0717	<u>0.0644</u>	0.0732	<b>0.0630</b>	0.0694	<u>0.0633</u>	0.0827	0.0891
Buddingtonite GDS85 D-206	<b>0.0706</b>	0.0805	0.0752	0.0841	<u>0.0729</u>	0.0742	0.0935	0.1012
Chalcedony CU91-6A	<b>0.1086</b>	0.1283	0.1301	0.1177	0.1388	0.1407	0.1362	<u>0.1110</u>
Dumortierite HS190.3B	<u>0.1051</u>	0.1105	0.1126	0.1067	<b>0.1045</b>	0.1096	0.1134	0.1229
Kaolin/Smect H89-FR-5 30K	0.0729	<b>0.0578</b>	0.0823	0.0705	0.0794	<u>0.0686</u>	0.0694	0.0892
Kaolin/Smect KLF508 85%K	<u>0.0890</u>	0.0978	0.0998	0.0993	0.1053	0.1095	<b>0.0876</b>	0.0969
Montmorillonite+Illite CM37	<u>0.0606</u>	0.0629	0.0729	0.0804	0.0847	0.0786	0.0824	<b>0.0592</b>
Muscovite GDS108	<u>0.1172</u>	<b>0.1163</b>	0.1276	0.1183	0.1445	0.1409	0.1420	0.1510
Nontronite NG-1.a	<b>0.0847</b>	0.0896	0.1015	<u>0.0880</u>	0.1189	0.0889	0.1148	0.0902
Pyrope WS474	<b>0.1462</b>	<u>0.1504</u>	0.1668	0.1580	0.1702	0.1781	0.1802	0.1709
Sphene HS189.3B	0.0918	0.1102	<b>0.0907</b>	0.0915	0.1364	0.0909	<u>0.0913</u>	0.1201
Mean	<b>0.0952</b>	0.0998	0.1047	0.1002	0.1125	0.1059	0.1139	0.1103



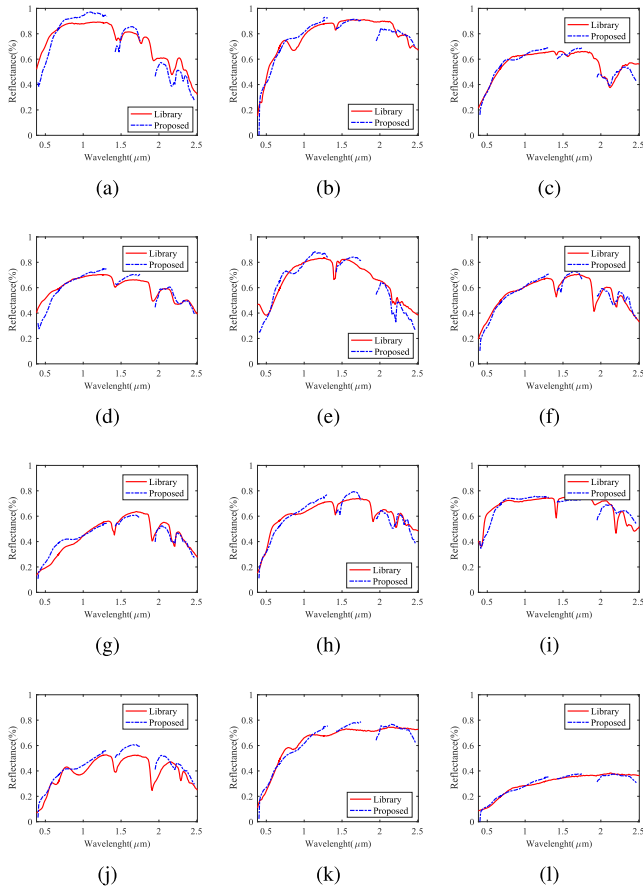


Fig. 15. Comparison of the USGS library spectra with the endmember signatures extracted by MPEC-NMF on the AVIRIS Cuprite dataset. (a) Alunite GDS82 Na82. (b) Andradite WS487. (c) Buddingtonite GDS85 D-206. (d) Chalcodony CU91-6 A. (e) Dumortierite HS190.3B. (f) Kaolin/Smect H89-FR-5 30 K. (g) Kaolin/Smect KLF508 85%K. (h) Montmorillonite + Illi CM37. (i) Muscovite GDS108. (j) Nontronite NG-1.a. (k) Pyrope WS474. (l) Sphene HS189.3B.

the USGS library. It can be obviously seen that the signatures extracted by proposed method have a high spatial similarity with the reference signatures, indicating that the method is efficient. The corresponding SAD values obtained for the different unmixing approaches are given in Table III. It can be observed in the table that most of the minerals estimated by the proposed method achieve the best performance and have a lower mean SAD value than other methods.

## VI. CONCLUSION AND FUTURE WORK

In this article, we attempted to unify the geometric *prior* and statistical *prior* of HSI into the same NMF unmixing model and impose constraints on both the endmember and the abundance matrix. A new unmixing scheme is proposed, namely multiple-*priors* ensemble constrained NMF (MPEC-NMF). The important *priors* of HSI: the endmember simplex volume, abundance sparsity, and abundance smoothness work together in the same model for the first time. The union of geometric-statistical *priors* and dual-matrix constraints effectively overcomes the limitations of the traditional single-matrix constraints method and improves the accuracy of the results. We propose

an efficient optimization algorithm for solving the proposed model by using variable splitting and augmented Lagrangian frameworks. Moreover, the experimental results on the synthetic and real-world datasets to demonstrate the effectiveness of the proposed method and achieved better performance than other state-of-the-art methods.

In future work, we will continue to explore the application of multiple *priors* ensemble constraints in HU issue, and the strategies and rules of multiple *priors* ensemble will be further attempted. In addition, we will continue to improve the performance of the algorithm, introducing a fast method instead of time-consuming operations.

## ACKNOWLEDGMENT

The authors would like to thank Dr. W. He for generously sharing the code of TV-RSNMF method.

## REFERENCES

- [1] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.
- [2] X. Xu *et al.*, "A new spectral-spatial sub-pixel mapping model for remotely sensed hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6763–6778, Nov. 2018.
- [3] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, 2015.
- [4] 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 Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [5] 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.
- [6] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [7] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, 1999.
- [8] 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 Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.
- [9] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. 1st Workshop Hyperspectral Image Signal Process.: Evolution Remote Sens.*, pp. 1–4, Aug. 2009.
- [10] 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 Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6245–6257, Oct. 2018.
- [11] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [12] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
- [13] 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.
- [14] 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.
- [15] 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.



- [16] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.
- [17] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 554–569, Apr. 2013.
- [18] P. Thouvenin, N. Dobigeon, and J. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.
- [19] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, 2016.
- [20] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
- [21] T. Uezato, M. Fauvel, and N. Dobigeon, "Hyperspectral unmixing with spectral variability using adaptive bundles and double sparsity," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3980–3992, Jun. 2019.
- [22] Z. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang, " $l_{1/2}$  regularization," *Sci. China Inf. Sci.*, vol. 53, pp. 1159–1169, 2010.
- [23] R. Ammanouil, A. Ferrari, and C. Richard, "A graph Laplacian regularization for hyperspectral data unmixing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 1637–1641.
- [24] 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.
- [25] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, Apr. 2019.
- [26] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Blind hyperspectral unmixing using total variation and  $l_q$  sparse regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6371–6384, Nov. 2016.
- [27] J. Yao, D. Meng, Q. Zhao, W. Cao, and Z. Xu, "Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2991–3006, Jun. 2019.
- [28] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.
- [29] K. Qu, W. Bao, and X. Shen, "Hyperspectral unmixing using weighted  $l_{1/2}$  sparse total variation regularized and volume prior constrained non-negative matrix factorization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2147–2150.
- [30] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [31] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [32] J. Fan and H. Peng, "Nonconcave penalized likelihood with a diverging number of parameters," *Ann. Statist.*, vol. 32, no. 3, pp. 928–961, 2004.
- [33] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted  $l_1$  minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
- [34] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [35] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [36] D. P. Bertsekas, *Nonlinear Programming: 2nd Edition*. Belmont, MA, USA: Athena Scientific, 1999.
- [37] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Softw.*, vol. 8, no. 1, pp. 43–71, 1982.
- [38] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, pp. 141–148, 1988.
- [39] 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.
- [40] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, no. 1, pp. 1457–1469, 2004.
- [41] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral *a priori* information," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 770–783, Feb. 2015.
- [42] 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.
- [43] 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.
- [44] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [45] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Tech. Rep. CS-94-125, Pittsburgh, PA, USA, 1994.
- [46] Y. Huang, H. Liu, and S. Zhou, "An efficient monotone projected Barzilai–Borwein method for nonnegative matrix factorization," *Appl. Math. Lett.*, vol. 45, pp. 12–17, 2015.
- [47] L. Han, M. Neumann, and U. Prasad, "Alternating projected Barzilai–Borwein methods for nonnegative matrix factorization," *Electron. Trans. Numer. Anal. ETNA*, vol. 36, pp. 54–82, 2010.
- [48] Y. Dai and L. Liao, "R-linear convergence of the Barzilai and Borwein gradient method," *IMA J. Numer. Anal.*, vol. 22, no. 1, pp. 1–10, 2002.
- [49] Y. Huang and H. Liu, "On the rate of convergence of projected Barzilai–Borwein methods," *Optim. Methods Softw.*, vol. 30, no. 4, pp. 880–892, 2015.
- [50] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.



**Kewen Qu** (Member, IEEE) received the B.Eng. degree in computer science from Henan University, Kaifeng, China, in 2008, and the M.Sc. degree in computer science from North Minzu University, Yinchuan, China, in 2012. He is currently working toward the Ph.D. degree with the School of Computer and Information, Hefei University of Technology, Hefei, China.

His current research interests include hyperspectral image processing, machine learning, and computer vision.



**Wenxing Bao** (Member, IEEE) received the B.Eng. degree in industrial automation from Xidian University, Xi'an, China, in 1993, the M.Sc. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2001, and the Ph.D. degree in electronic science and technology from Xi'an Jiaotong University, Xi'an, China, in 2006.

He is currently a Professor with North Minzu University, Yinchuan, China. His current research interests include digital image processing, remote sensing image classification and fusion.