

Spatial feature extraction non-negative tensor factorization for hyperspectral unmixing

Jin-Ju Wang, Ding-Cheng Wang, Ting-Zhu Huang*, Jie Huang*

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, PR China



ARTICLE INFO

Article history:

Received 7 March 2021

Revised 2 August 2021

Accepted 23 September 2021

Available online 22 October 2021

Keywords:

Hyperspectral unmixing

The feature layer

The sparse layer

Non-negative tensor factorization (NTF)

ABSTRACT

Estimating endmembers and corresponding abundances from mixed pixels are essential steps for hyperspectral unmixing. In hyperspectral unmixing, obtaining accurate unmixing results is difficult since less prior knowledge is available. Besides, the unmixing results are influenced by noise and highly correlated endmembers, so that the obtained abundance maps exist small values which are not present in the image. In this paper, we separate each abundance map into a feature layer and a sparse layer to protect the obtained abundance maps from the above-mentioned factors. The feature layer represents the main information of the abundance map. And the sparse layer contains outliers dominated by the above-mentioned factors. In particular, we design a feature extraction regularization to describe the feature layer and use a weighted ℓ_1 norm to describe the sparse layer. Then, under the framework of non-negative tensor factorization (NTF), we propose a novel unmixing algorithm named spatial feature extraction NTF (SFE-NTF) for hyperspectral unmixing. The proposed SFE-NTF is based on an augmented multiplicative algorithm. Experimental results on both synthetic and real hyperspectral data demonstrate that the proposed algorithm outperforms other state-of-the-art algorithms.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

Hyperspectral images (HSIs) contain wealthy spectral information, which are widely applied in various fields [1–4]. Due to the low spatial resolution of HSIs, different types of substances are mixed in a pixel, called mixed pixel. The mixed pixels limit further application of HSIs [5]. Thus, hyperspectral unmixing (HU), aims at finding the constituted substances (endmembers) and the corresponding fractions (abundances) of the mixed pixels, is critical for the application of HSIs.

The linear mixture model (LMM) is a classical way to account for the mixture process. LMM assumes that each pixel is a linear combination of the constituted endmembers, weighted by the corresponding abundances [6]. Usually, HU includes two steps: endmember extraction and abundance estimation. Endmember extraction algorithms mainly include pixel purity index [7], N-FINDR [8], vertex component analysis (VCA) [9], and so on. These algorithms extract the endmembers under the assumption that HSIs contain pure pixels, which may be incompatible with some practical cases. Abundance estimation algorithms mainly include three categories, which are geometry-based, statistics-based, and sparse-regression-based ones. Specifically, geometry-based ones assume that all pixels in the hyperspectral images are distributed in high-dimensional simplex space or in a positive convex cone region [10,11]. The statistics-based ones determine endmember set and abun-

* Corresponding authors.

E-mail addresses: tingzhuhuang@126.com (T.-Z. Huang), huangjie_uestc@uestc.edu.cn (J. Huang).

dance matrix by parameter estimation [12,13]. The sparse-regression-based ones mainly use the prior knowledge of spectral database as a dictionary to transform the separation into a sparse regression problem [14,15]. As one part in the statistics-based algorithms, non-negative matrix factorization (NMF) [16] and non-negative tensor factorization (NTF) [17] have been widely used in HU for its clear physical, statistical, geometric interpretation, and flexible modeling [18]. Here, we mainly discuss NMF and NTF for HU.

NMF decomposes a non-negative matrix into two non-negative matrices with a part-based representation [19,20]. Applied NMF to the matrix reshaped by a 3-D HSI, the obtained two non-negative matrices represent the endmember matrix and the abundance matrix, respectively. However, the decomposition from a 3-D HSI into a 2-D matrix causes information loss in the horizontal direction [17,21]. Thus, Qian et al. extend NMF into a tensor-based model and propose a matrix-vector NTF (MV-NTF) unmixing algorithm [17]. In MV-NTF, an HSI is decomposed into R component tensors, where each one is the outer product of a low-rank matrix (abundance map) and a vector (endmember). Due to its flexible modeling, NMF and MV-NTF can simultaneously estimate the endmembers and the abundances for HU. However, the non-convexity of NMF and MV-NTF makes that the solution is easy to fall into local minima.

To improve the performance of NMF and MV-NTF, various regularizers are taken into account. For example, sparsity constraint is incorporated into NMF to exploit the sparseness of the abundance matrix [22,23]. Sparsity constraint utilizes the fact that most pixels contain a few endmembers instead of all endmembers. Then, an ℓ_1 regularizer is applied on the abundance matrix to yield sparse solutions [22]. However, the ℓ_1 regularizer is insufficient to enforce the further sparsity of the abundance matrix sometimes. To provide a sparser solution, an $L_{1/2}$ regularizer is considered in NMF [24]. Indeed, the $L_{1/2}$ regularizer has been proved that can produce the best sparse solution than other L_q ($1/2 \leq q < 1$) regularizers in [24].

Besides, spatial information are also taken into account in HU to achieve more accurate results [23,25–27]. Typically, some researchers assume that both the constituted endmembers and their associated fractional abundances of the neighboring pixels should be similar, which is called piece-wise smoothness [28,29]. Then, the total variation (TV) regularization is applied on the abundance maps to describe the piece-wise smoothness [28,29]. Different from the TV regularization, a Laplacian graph regularization exploits the consistency of the abundances when the spectral signatures of neighboring pixels are similar [30]. Introducing the above regularizers into HU improves the unmixing performance. However, it's still hard to obtain accurate separation of the mixed spectral curves for the following aspects:

- 1) Since the neighboring pixels are not similar sometimes, the over strict smooth constraint may remove the edges in the image. In addition, the over strict smooth constraint may cause that the neighbors of edge pixels are assigned some values which are not in the image.
- 2) The Laplacian graph regularization is dependent on the similarity of spectral curves in HSI, which is vulnerable to noise.
- 3) Some endmembers in HSIs are highly correlated. Geometrically speaking, it means that one endmember is similar to another endmember or the affine hull spanned by other composed endmembers in HSIs [11]. In this case, it is difficult to estimate highly correlated endmembers from the mixed pixels.

The above factors seriously influence the decomposition of mixed pixels since less prior knowledge is available. In this case, the estimated abundance maps may contain two parts: one is the main information in the true abundance maps, and the other is composed of outliers generated by the above-mentioned factors. To remove these outliers from the estimated abundance maps, we decompose each abundance map into the feature layer and the sparse layer. The feature layer is the main information of the abundance map and the sparse layer represents the outliers. In detail, we give an example of the Jasper Ridge data in Fig. 1. As shown in Fig. 1, endmembers distribute densely in the feature layers, and distribute rarely in the sparse layers. In this paper, we adopt two different constraints to exploit these two layers.

Recall that MV-NTF avoids information loss and improves the unmixing results, in this paper, we propose a novel NTF based unmixing algorithm named spatial feature extraction non-negative tensor factorization (SFE-NTF) for hyperspectral unmixing. As the framework shown in Fig. 1, we first divide the HSI into endmembers and the corresponding abundance maps, and decompose each abundance map into two layers: feature layer and sparse layer. In SFE-NTF, the feature layer is exploited by a spatial feature extraction regularization, and the sparse layer is exploited by the weighted ℓ_1 norm. The proposed SFE-NTF algorithm is based on the augmented multiplicative algorithm. Experiments on both synthetic and real data have shown the advantages of our algorithm over several state-of-the-art approaches. In summary, the main contributions for this paper are:

1. We decompose each abundance map into the feature layer and the sparse layer to protect the unmixing results from noise and the highly correlated endmembers. The feature layer represents the main information of the abundance maps, while the sparse layer represents the outliers.
2. We introduce a spatial feature extraction regularization into the feature layer to protect it from factors like noise. In addition, the sparse layer appears sparser than the original abundance map, thus we apply the weighted ℓ_1 norm on the sparse layer to remove the outliers.

The rest of this paper is organized as follows. Section II briefly introduces the NMF and the NTF model. Section III shows the details of the proposed model, the associated update rules. The experimental results on synthetic data are presented and analyzed in Section IV. In Section V, real-data experiments are conducted. Finally, conclusions and future work are given in Section VI.

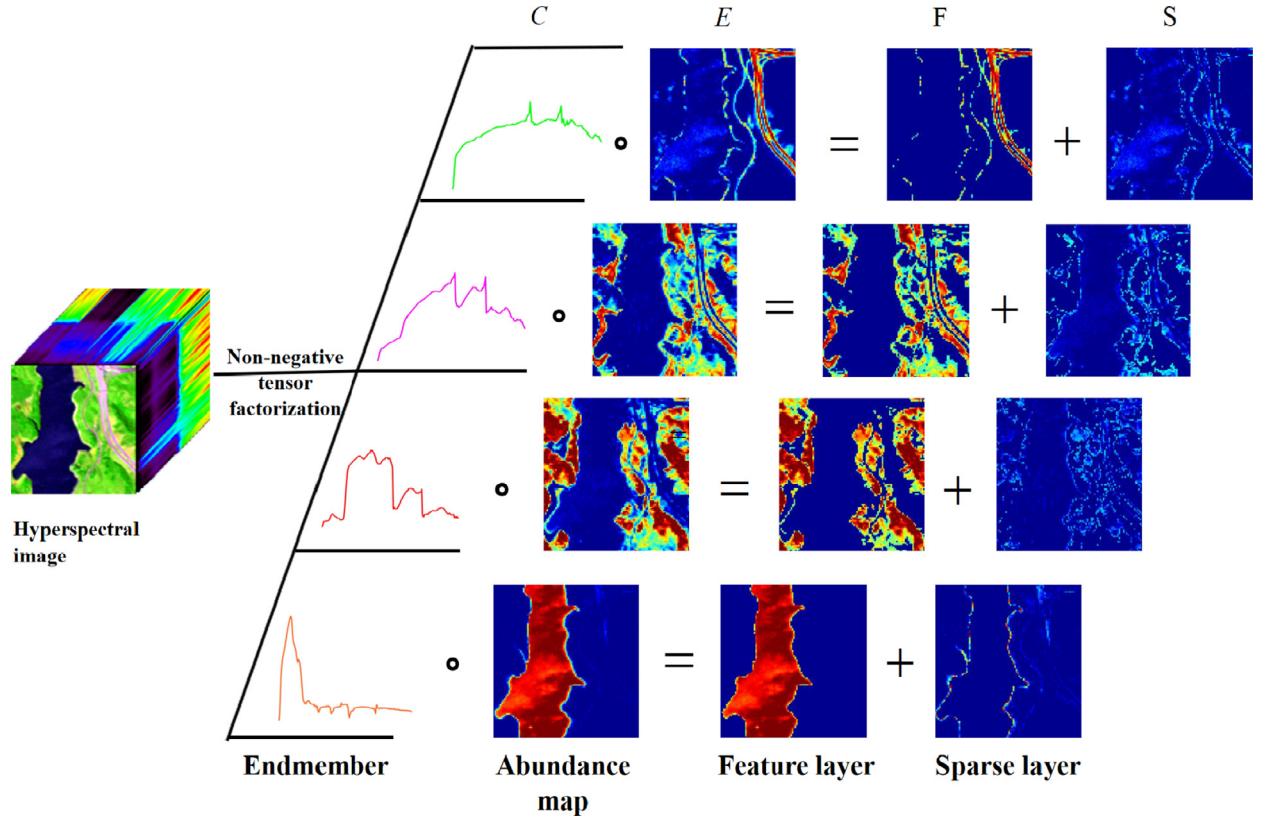


Fig. 1. The flowchart of the proposed algorithm.

2. Notations and basic unmixing models

This section first introduces some notations. We then review the framework of NMF and MV-NTF.

2.1. Notations

In this paper, we have the following representations:

scalars	lowercase letters	y
vectors	boldface lowercase letters	$\mathbf{y} \in \mathbb{R}^I$
matrices	boldface capital letters	$\mathbf{Y} \in \mathbb{R}^{I \times J}$
tensors	boldface Euler script letters	$\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$

Definition of the Outer Product: Given K vectors $\mathbf{b}^{(i)} \in \mathbb{R}^{N_i}, i = 1, \dots, K$, the outer product between them is a tensor $\mathcal{X} = \mathbf{b}^{(1)} \circ \mathbf{b}^{(2)} \circ \dots \circ \mathbf{b}^{(K)} \in \mathbb{R}^{N_1 \times N_2 \times \dots \times N_K}$ [31]. Each element of the tensor \mathcal{X} is the product of the corresponding vector elements:

$$x_{i_1 i_2 \dots i_K} = b_{i_1}^{(1)} \times b_{i_2}^{(2)} \dots b_{i_K}^{(K)} \quad \text{for all } 1 \leq i_n \leq N_i$$

Definition of the Kronecker Product and Khatri-Rao Product:

Given two matrices $\mathbf{A} \in \mathbb{R}^{I \times J}, \mathbf{B} \in \mathbb{R}^{K \times L}$, their Kronecker product, $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{(IK) \times (JL)}$, is defined by

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B} &= \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \dots & a_{IJ}\mathbf{B} \end{pmatrix} \\ &= [\mathbf{a}_1 \otimes \mathbf{b}_1 \ \mathbf{a}_1 \otimes \mathbf{b}_2 \ \mathbf{a}_1 \otimes \mathbf{b}_3 \dots \mathbf{a}_J \otimes \mathbf{b}_{L-1} \ \mathbf{a}_J \otimes \mathbf{b}_L], \end{aligned}$$

where \mathbf{a}_j ($j = 1, 2, \dots, J$) denotes the j -th column in \mathbf{A} and a_{ij} denotes the (i, j) -th element of \mathbf{A} [31]. A similar denotation is used for \mathbf{B} .

When \mathbf{A} and \mathbf{B} have the same number of columns, i.e., $J = L$, their Khatri-Rao product $\mathbf{A} \odot \mathbf{B} \in \mathbb{R}^{(IK) \times J}$, is denoted by

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \ \mathbf{a}_2 \otimes \mathbf{b}_2 \ \cdots \ \mathbf{a}_J \otimes \mathbf{b}_J].$$

When $\mathbf{A} = [\mathbf{A}_1 \cdots \mathbf{A}_R]$ and $\mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_R]$ are two block matrices with the same number of submatrices, the generalized Khatri-Rao product for partitioned matrices is

$$\mathbf{A} \bar{\odot} \mathbf{B} = [\mathbf{A}_1 \odot \mathbf{B}_1 \ \mathbf{A}_2 \odot \mathbf{B}_2 \ \cdots \ \mathbf{A}_R \odot \mathbf{B}_R].$$

2.2. Basic unmixing models

1) NMF: Given an HSI $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ with $I \times J$ pixels and K bands, we first reshape \mathcal{Y} into a matrix $\mathbf{Y} \in \mathbb{R}^{K \times IJ}$. Then, the unmixing model based on the LMM is written as

$$\mathbf{Y} = \mathbf{CX} + \mathbf{N}, \quad (1)$$

where $\mathbf{C} \in \mathbb{R}^{K \times R}$ is the endmember matrix with R endmembers, and $\mathbf{X} \in \mathbb{R}^{R \times IJ}$ is the abundance matrix, whose r -th row represents the abundances of the r -th endmember for IJ pixels. Meanwhile, \mathbf{N} represents the additive white Gaussian noise. According to the physical meanings of abundances, the abundance matrix usually meets two constraints: the abundance non-negativity constraint (ANC, i.e., $\mathbf{X} \geq \mathbf{0}$) and the abundance sum-to-one constraint (ASC, i.e., $\mathbf{1}_R^T \mathbf{X} = \mathbf{1}_{IJ}^T$) [6]. Then, the loss function of NMF is

$$\min_{\mathbf{C}, \mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{CX}\|_F^2, \quad \text{s.t. } \mathbf{C}, \mathbf{X} \geq \mathbf{0}, \quad (2)$$

2) MV-NTF: Note that reshaping an HSI into a 2-D matrix inevitably causes spatial information loss in the horizontal direction. Then, in [17], Qian et al. consider HSIs as 3-D tensors and propose MV-NTF. In MV-NTF, an HSI is considered as the sum of the outer product of the endmembers and their corresponding abundance maps, which is consistent with the LMM. The MV-NTF is described as

$$\begin{aligned} \mathcal{Y} &= \sum_{r=1}^R \mathbf{E}_r \circ \mathbf{c}_r + \mathcal{N} = \sum_{r=1}^R \mathbf{A}_r \mathbf{B}_r^T \circ \mathbf{c}_r + \mathcal{N}, \\ \text{s.t. } \mathbf{E}_1 + \cdots + \mathbf{E}_R &= \mathbf{1}_{I \times J}, \end{aligned} \quad (3)$$

where $\mathbf{1}_{I \times J} \in \mathbb{R}^{I \times J}$ is the matrix with all elements are 1. And $\mathbf{E}_r \in \mathbb{R}^{I \times J}$ represents the abundance map for the r -th endmember \mathbf{c}_r in \mathbf{C} . To investigate the low-rankness of the abundance map, \mathbf{E}_r is represented by the product of two low-rank matrices $\mathbf{A}_r \in \mathbb{R}^{I \times L_r}$ and $\mathbf{B}_r \in \mathbb{R}^{J \times L_r}$. Decomposing the HSI in three directions, we write the matrix representation of (3) as

$$\mathbf{Y}_{(1)} = [(\mathbf{A}_1 \odot \mathbf{B}_1) \mathbf{1}_{L_1} \cdots (\mathbf{A}_R \odot \mathbf{B}_R) \mathbf{1}_{L_R}] \mathbf{C}^T,$$

$$\mathbf{Y}_{(2)} = (\mathbf{B} \bar{\odot} \mathbf{C}) \mathbf{A}^T,$$

$$\mathbf{Y}_{(3)} = (\mathbf{C} \bar{\odot} \mathbf{A}) \mathbf{B}^T,$$

where $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_R]$, $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_R]$. And $\mathbf{1}_{L_r} \in \mathbb{R}^{L_r}$ is a vector with all ones.

Then, the loss function of MV-NTF is

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \frac{1}{2} \|\mathcal{Y} - \sum_{r=1}^R \mathbf{A}_r \mathbf{B}_r^T \circ \mathbf{c}_r\|_F^2 + \frac{\delta}{2} \|\mathbf{1}_{I \times J} - \mathbf{AB}^T\|_F^2, \quad (4)$$

where $\delta \geq 0$ is a parameter for the sum-to-one constraint.

3. Proposed algorithm

Due to the effect of noise and the highly correlated endmembers, the estimated abundance maps may exist some small abundances which are not in the image. In this case, the estimated abundance maps may contain two parts: one is the main information of the abundance maps and the other is generated by the factors like noise. Thus, we separate each abundance map into the feature layer and the sparse layer to obtain accurate abundance maps. To exploit the property of these two layers, we introduce two regularizers. Then, considering the NTF's natural representation for HSIs, we write our objective function as

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{F}, \mathbf{S}, \mathbf{C}} \quad & \frac{1}{2} \|\mathcal{Y} - \sum_{r=1}^R \mathbf{E}_r \circ \mathbf{c}_r\|_F^2 + \frac{\delta}{2} \|\mathbf{1}_{I \times J} - \mathbf{ED}\|_F^2 \\ & + \lambda f(\mathbf{F}) + \beta g(\mathbf{S}) \\ \text{s.t. } \mathbf{E}, \mathbf{C} & \geq \mathbf{0}, \mathbf{E} = \mathbf{F} + \mathbf{S}, \end{aligned} \quad (5)$$

where $\mathbf{E} = [\mathbf{E}_1, \dots, \mathbf{E}_R]$, $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R]$, $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_R]$, $\mathbf{D}_i \in \mathbb{R}^{I \times I}$, ($i = 1, 2, \dots, R$), are identity matrices. And $f(\mathbf{F})$ is the spatial feature extraction regularization applied on the feature layer, $g(\mathbf{S})$ is the constraint applied on the sparse layer, λ , and β are the parameters for $f(\mathbf{F})$ and $g(\mathbf{S})$, respectively. In the following, we first introduce the spatial feature extraction regularization and the weighted sparsity constraint. Then, we present the proposed model and the corresponding algorithm.

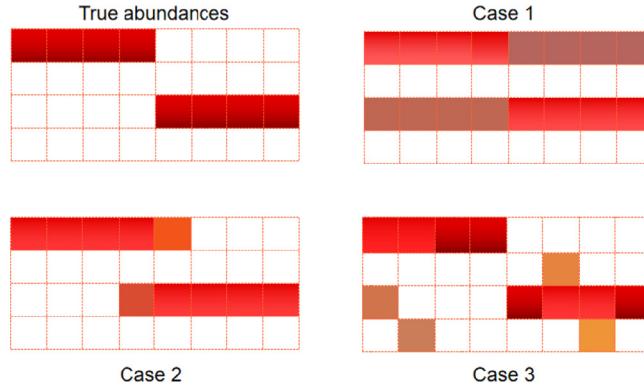


Fig. 2. Different cases that may make the separation of mixed pixels inaccurately. Case 1: the highly correlated endmembers; Case 2: the over-strict smooth constraint; Case 3: noisy environment;.

3.1. Spatial feature extraction regularization

In practical cases, there are many factors that make estimating abundances difficult. As shown in Fig. 2, we consider a sub-image of HSI, which covers two endmembers: endmembers 1 and 3. In this sub-image, endmember 1 is distributed in the first four pixels, and endmember 3 is distributed in the rest pixels. Here, we mainly discuss three different cases in the unmixing process that may make the separation of mixed pixels inaccurately.

Case 1: Due to the high correlation of these two endmembers, the spectral curves of endmember 1 may be identified by endmember 3. Then, the abundances of endmember 1 may be degraded. The same situation may also occur to the abundances of the endmember 3.

Case 2: Due to the strict smooth constraint applied on the abundance maps, edge regions may be removed and the neighbors of the edge pixels may be assigned some abundances.

Case 3: Due to existence of noise in the HSI, the estimated abundance maps may exist small abundances generated by noise.

Thus, we introduce a spatial feature extraction regularization on the feature layer, which aims at protecting the regions belonging to the image from the above cases. Recall that the feature layer is the main information of the abundance maps, thus we use Ostu's segmentation algorithm on each abundance map to capture the feature layer. Ostu's algorithm is a widely used technique to segment an image for further processing, such as feature analysis and quantification [32]. It searches for a threshold that minimizes the intra-class variances across the whole range of the pixel values and achieves good results. When the histogram of the original image has two distinct peaks, one belongs to the background, and the other belongs to the foreground of the signal. According to the definition of the feature layer and the sparse layer, applied Ostu's algorithm on each abundance map, the background and the foreground represent the feature layer and the sparse layer, respectively.

Then, we define this spatial feature extraction regularization as follows

$$f(\mathbf{F}) = \|\mathbf{F} - \hat{\mathbf{F}}\|_F, \quad (6)$$

where $\hat{\mathbf{F}}$ is the foreground obtained by applying Ostu's algorithm on \mathbf{E} .

3.2. Weighted sparsity constraint

In practice, endmembers are not distributed in the whole scene, which means that abundance maps cover a lot of zero components[24]. In other words, abundance map $\mathbf{X} \in \mathbb{R}^{I \times J}$ is sparse, which is investigated by

$$\min_{\mathbf{X}} \|\mathbf{X}\|_0, \quad (7)$$

where $\|\mathbf{X}\|_0$ denotes the number of nonzero components of \mathbf{X} . However, the ℓ_0 minimization problem is non-convex, and is difficult to solve [33]. Thus, the ℓ_0 minimization problem is usually relaxed to the following ℓ_1 minimization problem [34,35]:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_1 = \sum_{i,j} |x_{i,j}|, \quad (8)$$

where $x_{i,j}$ denotes the element (i, j) th of \mathbf{X} . Though the ℓ_1 minimization problem is convex and easy to solve, it is hard to guarantee further sparsity when the ASC is applied. To enhance the sparsity, Candes et al. weight the ℓ_1 norm and iteratively update the weights [36]. In addition, Candes et al. prove this reweighted strategy works. The reweighted ℓ_1 minimization problem is as follows:

$$\min_{\mathbf{X}} \|\mathbf{W} \cdot \mathbf{X}\|_1, \quad (9)$$

where $\mathbf{W} \in \mathbb{R}^{I \times J}$ is a weighting matrix, and the operator \cdot denotes the element-wise multiplication.

Recall that the sparse layer represents outliers generated by interference like noise, which are not distributed in the whole image. Thus, the sparse layer is sparse, and we apply the above reweighted ℓ_1 norm on the sparse layer

$$g(\mathbf{S}) = \|\mathbf{W} \cdot * \mathbf{S}\|_1. \quad (10)$$

3.3. Proposed model

Then, we formulate the proposed model by introducing the spatial feature extraction regularization and the weighted sparsity constraint into the NTF framework.

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{F}, \mathbf{S}, \mathbf{C}} \quad & \frac{1}{2} \|\mathcal{Y} - \sum_{r=1}^R \mathbf{E}_r \circ \mathbf{c}_r\|_F^2 + \frac{\delta}{2} \|\mathbf{1}_{I \times J} - \mathbf{ED}\|_F^2 \\ & + \frac{\lambda}{2} \sum_{r=1}^R \|\mathbf{F}_r - \hat{\mathbf{F}}_r\|_F^2 + \beta \sum_{r=1}^R \|\mathbf{S}_r\|_{w,1} \\ \text{s.t.} \quad & \mathbf{E}, \mathbf{C} \geq \mathbf{0}, \mathbf{E} = \mathbf{F} + \mathbf{S}, \end{aligned} \quad (11)$$

where $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_R]$, $\mathbf{F}_i \in \mathbb{R}^{I \times J}$, ($i = 1, \dots, R$) is the feature layer of \mathbf{E}_r . And $\mathbf{S} = [\mathbf{S}_1, \dots, \mathbf{S}_R]$ denotes the sparse layer, where \mathbf{S}_r is the sparse layer of \mathbf{E}_r .

We now solve (11) by the augmented multiplicative algorithm [16]. And the objective function is modified as

$$\begin{aligned} \mathcal{L}(\mathbf{E}, \mathbf{C}, \mathbf{F}, \mathbf{S}) = & \frac{1}{2} \|\mathcal{Y} - \sum_{r=1}^R \mathbf{E}_r \circ \mathbf{c}_r\|_F^2 + \frac{\delta}{2} \|\mathbf{1}_{I \times J} - \mathbf{ED}\|_F^2 \\ & + \frac{\lambda}{2} \sum_{r=1}^R \|\mathbf{F}_r - \hat{\mathbf{F}}_r\|_F^2 + \beta \sum_{r=1}^R \|\mathbf{W}_r \cdot * \mathbf{S}_r\|_1 \\ & + \frac{\mu}{2} \|\mathbf{E} - \mathbf{F} - \mathbf{S}\|_F^2 + \text{Tr}(\Gamma \mathbf{E}) + \text{Tr}(\Phi \mathbf{C}), \end{aligned} \quad (12)$$

where Γ, Φ are Lagrange multipliers for each non-negative variable in (12), \mathbf{W}_r is the weighting matrix for \mathbf{S}_r , and $\mu > 0$ is a penalty parameter. Considering that all variables in (12) are coupled, we alternatively solve one subproblem of the Lagrange function \mathcal{L} with other variables fixed.

For the \mathbf{E} -subproblem, remaining the terms corresponding to \mathbf{E} in (12), we obtain

$$\begin{aligned} \mathbf{E}^{(t+1)} = & \underset{\mathbf{E}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y}_{(1)} - \mathbf{M}_1 \mathbf{E}^T\|_F^2 + \frac{\delta}{2} \|\mathbf{1}_{I \times J} - \mathbf{ED}\|_F^2 \\ & + \frac{\mu}{2} \|\mathbf{E} - \mathbf{F} - \mathbf{S}\|_F^2, \end{aligned} \quad (13)$$

where $\mathbf{M}_1 = \mathbf{D} \odot \mathbf{C}$. Considering the convexity of the problem (13) and the Karush-Kuhn-Tucker (KKT) conditions, when the problem (13) reaches its optimal value, we have

$$(\mathbf{Y}_{(1)}^T - \mathbf{EM}_1^T) \mathbf{M}_1 + \delta (\mathbf{1}_{I \times J} - \mathbf{ED}) \mathbf{D}^T + \mu (\mathbf{E} - \mathbf{F} - \mathbf{S}) = \Gamma, \quad (14)$$

$$\Gamma \cdot * \mathbf{E} = \mathbf{0}. \quad (15)$$

Combining (14) and (15), we have the update rule for \mathbf{E}

$$\begin{aligned} \mathbf{E}^{(t+1)} = & \mathbf{E}^{(t)} \cdot * (\mathbf{Y}_{(1)}^T \mathbf{M}_1 + \delta \mathbf{1}_{I \times J} \mathbf{D}^T + \mu (\mathbf{F}^{(t)} + \mathbf{S}^{(t)})) \\ & ./ (\mathbf{E}^{(t)} \mathbf{D}^T \mathbf{D} + \mathbf{E}^{(t)} \mathbf{M}_1^T \mathbf{M}_1 + \mu \mathbf{E}^{(t)}). \end{aligned} \quad (16)$$

For the \mathbf{C} -subproblem, finding the terms that including \mathbf{C} in (12), we get

$$\mathbf{C}^{(t+1)} = \underset{\mathbf{C}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y}_{(3)} - \mathbf{M}_2 \mathbf{C}^T\|_F^2 + \text{Tr}(\Phi \mathbf{C}), \quad (17)$$

where $\mathbf{M}_2 = [(\mathbf{E}_1 \odot \mathbf{D}_1) \mathbf{1}_L \cdots (\mathbf{E}_R \odot \mathbf{D}_R) \mathbf{1}_L]$. Similar strategy to the \mathbf{E} -subproblem, the solution of the \mathbf{C} -subproblem is

$$\mathbf{C}^{(t+1)} = \mathbf{C}^{(t)} \cdot * (\mathbf{Y}_{(3)}^T \mathbf{M}_2) ./ (\mathbf{C}^{(t)} \mathbf{M}_2^T \mathbf{M}_2). \quad (18)$$

Decomposing the \mathbf{F} -subproblem into R subproblems, we have

$$\mathbf{F}_r^{(t+1)} = \underset{\mathbf{F}_r}{\operatorname{argmin}} \frac{\lambda}{2} \|\mathbf{F}_r - \hat{\mathbf{F}}_r\|_F^2 + \frac{\mu}{2} \|\mathbf{F}_r - (\mathbf{E}_r - \mathbf{S}_r)\|_F^2. \quad (19)$$

In the same way, we obtain the solution of the \mathbf{F}_r -subproblem

$$\mathbf{F}_r^{(t+1)} = \mathbf{F}_r^{(t)} \cdot * (\lambda \hat{\mathbf{F}}_r^{(t+1)} + \mu (\mathbf{E}_r^{(t+1)} - \mathbf{S}_r^{(t)})) ./ ((\lambda + \mu) \mathbf{F}_r^{(t)}), \quad (20)$$

where $\hat{\mathbf{F}}_r^{(t+1)}$ is the foreground obtained by applying Ostu's algorithm on $\mathbf{E}_r^{(t+1)}$.

To solve the \mathbf{S} -subproblem, we divide the \mathbf{S} -subproblem into R subproblems:

$$\mathbf{S}_r^{(t+1)} = \underset{\mathbf{S}_r}{\operatorname{argmin}} \beta \|\mathbf{W}_r \odot \mathbf{S}_r\|_1 + \frac{\mu}{2} \|\mathbf{S}_r - (\mathbf{E}_r - \mathbf{F}_r)\|_F^2. \quad (21)$$

Then, from [37,38], the solution of the \mathbf{S}_r -subproblem is

$$\mathbf{S}_r^{(t+1)} = \text{soft}_{\mathbf{W}_r^{(t+1)}}(\mathbf{E}_r^{(t+1)} - \mathbf{F}_r^{(t+1)}, \frac{\beta}{\mu}), \quad (22)$$

where $\text{soft}(u, \tau)$ is a component-wise application of the soft-threshold function [37,38]. And it is defined as

$$\text{soft}(u, \tau) = \operatorname{sign}(u) \max\{|u| - \tau, 0\}.$$

In (22), \mathbf{W}_r is updated element-wise by

$$\mathbf{W}_r^{(t+1)} = \frac{1}{\mathbf{S}_r^{(t)} + \epsilon}, \quad (23)$$

where $\epsilon > 0$ is added to avoid the singularity.

To make it more clear, we summarize the proposed algorithm, named SFE-NTF, in [Algorithm 1](#).

Algorithm 1 SFE-NTF.

Input: An HSI cube $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$;

The parameters δ, λ, β , and μ .

The number of endmembers R .

Initialization: $\mathbf{E}^{(0)}, \mathbf{C}^{(0)}, \mathbf{F}^{(0)}, \mathbf{S}^{(0)}$, and $\mathbf{W}^{(0)}$.

Repeat:

 Update $\mathbf{E}^{(t+1)}$ by (16)

 Update $\mathbf{C}^{(t+1)}$ by (18)

 Update $\mathbf{F}_r^{(t+1)}$ by (20)

 Update $\mathbf{S}_r^{(t+1)}$ by (22)

 Update $\mathbf{W}^{(t+1)}$ by (23)

Until termination condition is met.

Output: \mathbf{E} and \mathbf{C} .

4. Synthetic data experiments

This section illustrates the unmixing performance of the proposed algorithm by experimenting on two synthetic data cubes. Besides, we also compare the proposed algorithm with some relative works, $L_{1/2}$ -NMF [24], GLNMF [30], TV-RSNMF [28], SSWNMF [39], and MV-NTF-TV[21]. All tests are implemented on the platform of Windows 10 and MATLAB (R2019a) with an Intel Core i9-9900K, 3.60GHz, and 32GB RAM.

To quantitatively illustrate the unmixing performance, we select two popular criteria: spectral angular distance (SAD) and root mean square error (RMSE). Specifically, SAD_k evaluates the spectral angle between the true endmember signature $\hat{\mathbf{c}}_k$ and the estimated endmember signature \mathbf{c}_k , which is defined as

$$SAD_k = \arccos \left(\frac{\hat{\mathbf{c}}_k^T \mathbf{c}_k}{\|\hat{\mathbf{c}}_k\|_2 \|\mathbf{c}_k\|_2} \right).$$

And $RMSE_k$ measures the difference between the k -th true abundance map $\hat{\mathbf{E}}_k$ and the estimated abundance map \mathbf{E}_k , which is defined as

$$RMSE_k = \left(\frac{1}{N} \|\mathbf{E}_k - \hat{\mathbf{E}}_k\|_F^2 \right)^{\frac{1}{2}},$$

where $N = I \times J$ is the number of pixels.

In the experiments, we run NMF for 100 times to initialize $\mathbf{E}^{(0)}$ and $\mathbf{C}^{(0)}$, similarly to the other comparing algorithms. In addition, we set $\mathbf{F}^{(0)} = \mathbf{E}^{(0)}$, $\mathbf{S}^{(0)} = \mathbf{0}$, and $\mathbf{W}^{(0)} = \mathbf{0}$. Then, we show how to choose the parameters in SFE-NTF. For the data whose number of endmembers R is unknown, we use the corresponding estimation method, HySime [40], which is widely used in many unmixing works [21,41].

The selections of the optimal parameters are as follows

$$\lambda, \beta \in \{0, 0.01, 0.05, 0.1, 0.3, 0.5, 1, 3, 5, 10\},$$

$$\delta \in \{0.05, 0.1, 0.5, 1, 5, 10, 20, 30\},$$

$$\mu \in \{0.01, 0.1, 1, 5, 10, 20\}.$$

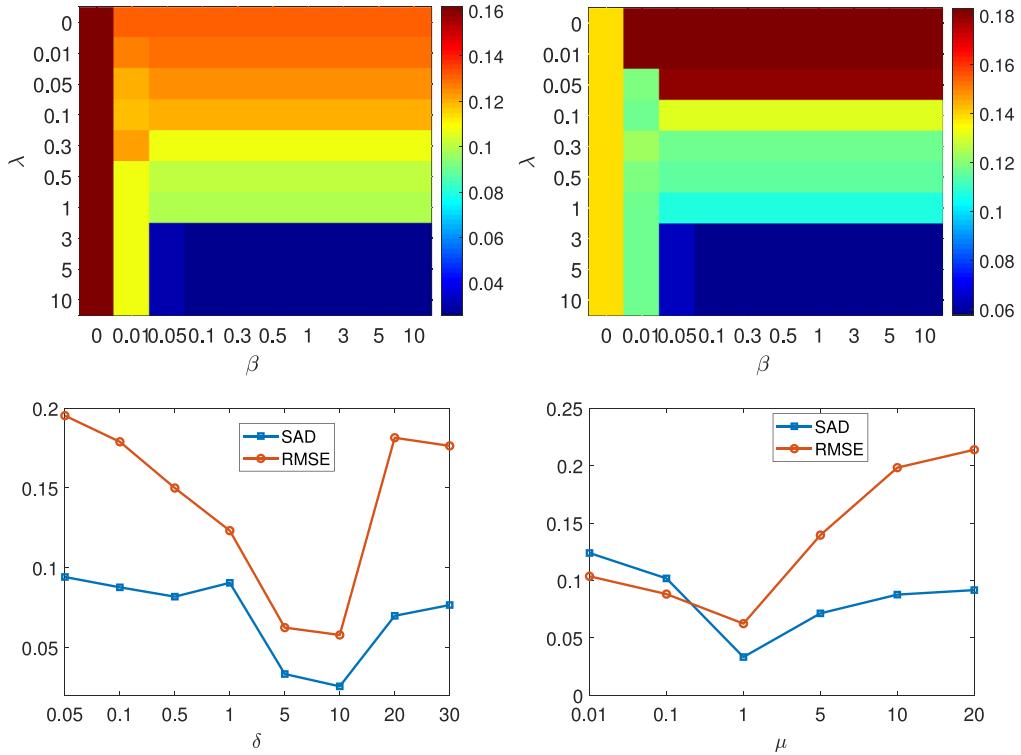


Fig. 3. Variation of mean SAD and mean RMSE of the proposed algorithm against parameters δ , λ , β , and μ on DC1.

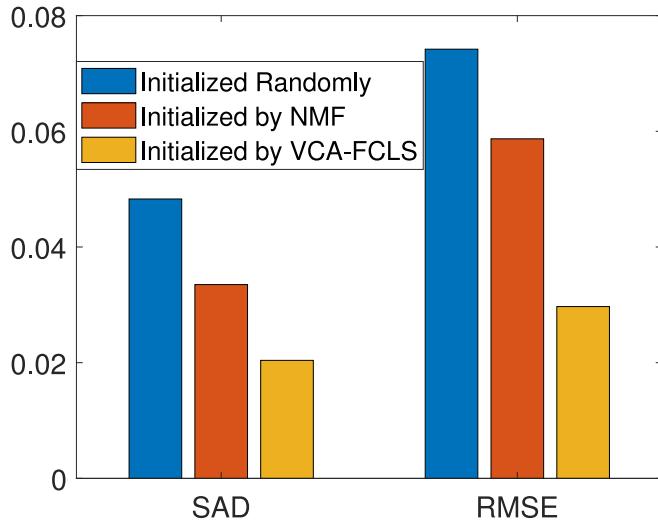


Fig. 4. Values of SAD and RMSE obtained by SFE-NTF with different initialization strategies on DC1.

4.1. Synthetic data cubes

We compare the unmixing performance of the comparing algorithms on the following simulated data cubes.

1) Simulated Data Cube 1 (DC1): DC1 contains 128×128 pixels with 431 bands, which has been widely used in many unmixing algorithms [42]. By LMM, five endmembers are selected from the U.S. Geological Survey (USGS) library to generate DC1, and the corresponding abundance maps are shown in Fig. 6.

2) Simulated Data Cube 2 (DC2): DC2 contains 100×100 pixels with 100 bands, which has been adopted in [44,45]. According to LMM, nine endmembers are selected from a library of 262 spectral signatures found on satellites, which is from the National Aeronautics and Space Administration Johnson Space Center (NASA JSC) [46]. And the corresponding abundance

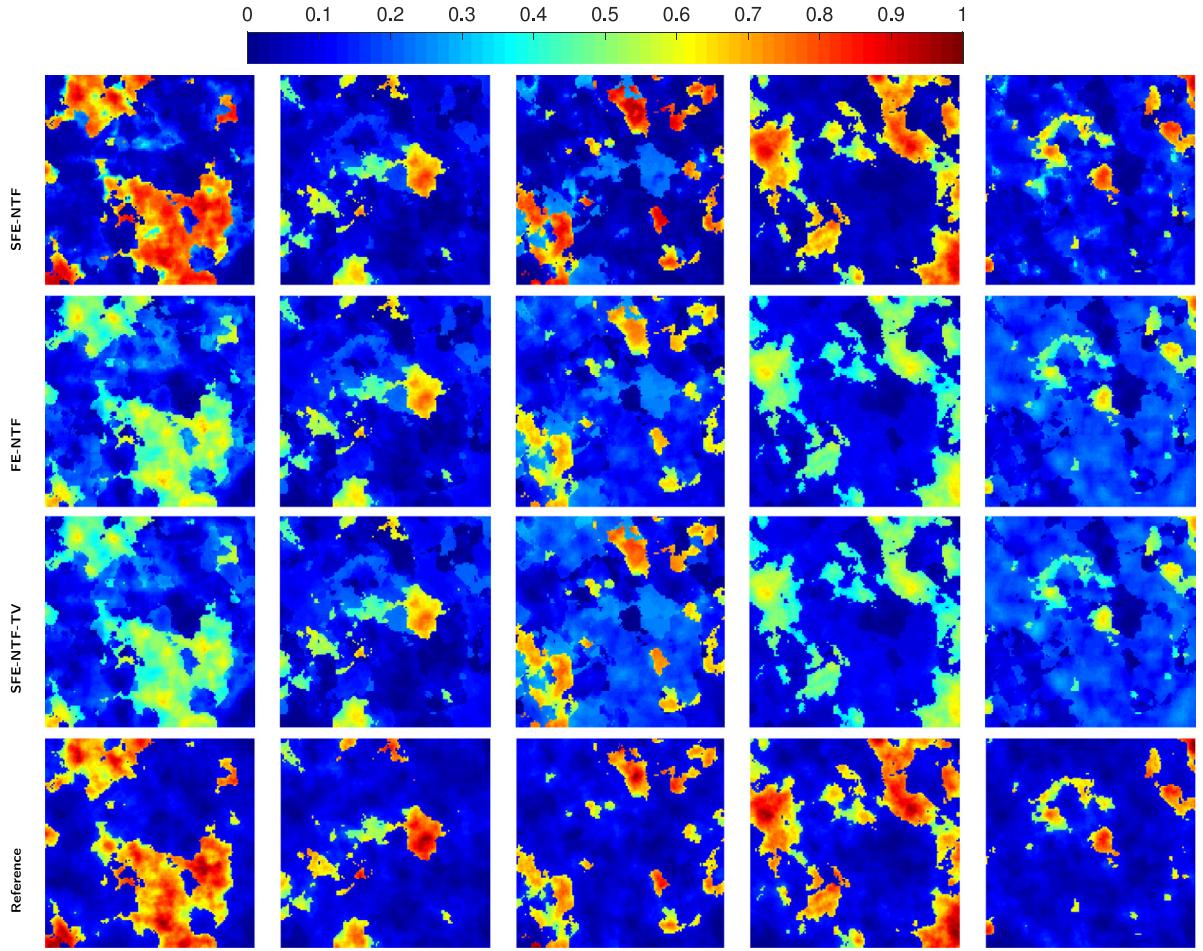


Fig. 5. Estimated abundance maps by SFE-NTF, FE-NTF, and SFE-NTF-TV for endmembers 1 to 5 on DC1.

maps are shown in [Figure 10](#). Then, the generated data are degraded by the noise with the signal-to-noise ratio (SNR) of 30 dB.

4.2. Experimental settings

Here, we discuss the parameters and the initial strategy in SFE-NTF.

1) Parameters analysis: In SFE-NTF, there are four parameters: δ , λ , β , and μ to influence the unmixing performance. We first fix $\delta = 5$ and $\mu = 1$ to test the sensitivity of the parameters λ and β . Then, we select λ and β from the sequence $\{0, 0.01, 0.05, 0.1, 0.3, 0.5, 1, 3, 5, 10\}$ and calculate values of SAD and RMSE as a function of parameters λ and β on DC1 in [Fig. 3](#). In the first row of [Fig. 3](#), the color of the region is closer to blue, the better the result by the corresponding λ and β .

From [Fig. 3](#), when $\lambda = 0, \beta = 0$, values of SAD and RMSE are larger than other choices of λ and β . When $\lambda \in \{3, 5, 10\}$ and $\beta \in \{0.1, 0.3, 0.5, 1, 3, 5, 10\}$, values of SAD and RMSE are stable and better than other choices of λ and β .

For the parameters δ and μ , we first fix $\lambda = 5, \beta = 1$, and $\mu = 1$ and plot values of SAD and values of RMSE as a function of $\delta \in \{0.05, 0.1, 0.5, 1, 5, 10, 20, 30\}$ in [Fig. 3](#). According to [Fig. 3](#), we obtain the lowest SAD and RMSE when $\delta = 10$.

Then, setting $\lambda = 5, \beta = 1$, and $\delta = 10$, we plot SAD and RMSE when $\mu \in \{0.01, 0.1, 1, 5, 10, 20\}$ in [Fig. 3](#). From [Fig. 3](#), we get the best SAD and RMSE when $\mu = 1$.

2) Influence of initialization: This experiment analyzes the unmixing performance of SFE-NTF with three different initial strategies: random initialization, NMF initialization, and VCA-FCLS initialization. [Fig. 4](#) shows values of SAD and RMSE obtained by SFE-NTF with different initialization strategies on DC1. From [Fig. 4](#), our algorithm obtains the best result with VCA-FCLS initialization. Meanwhile, when we adopt the other two initial strategies, SFE-NTF also obtains promising results.

3) Influence of Ostu's algorithm: In SFE-NTF, we utilize Ostu's algorithm to explore the information of the feature layer. To illustrate the influence of Ostu's algorithm in the decomposition strategy, we compare SFE-NTF with the following models: 1) SFE-NTF with $\mathbf{E} = \mathbf{F}$, called FE-NTF; 2) SFE-NTF with TV regularization to constrain the feature layer instead of the spatial

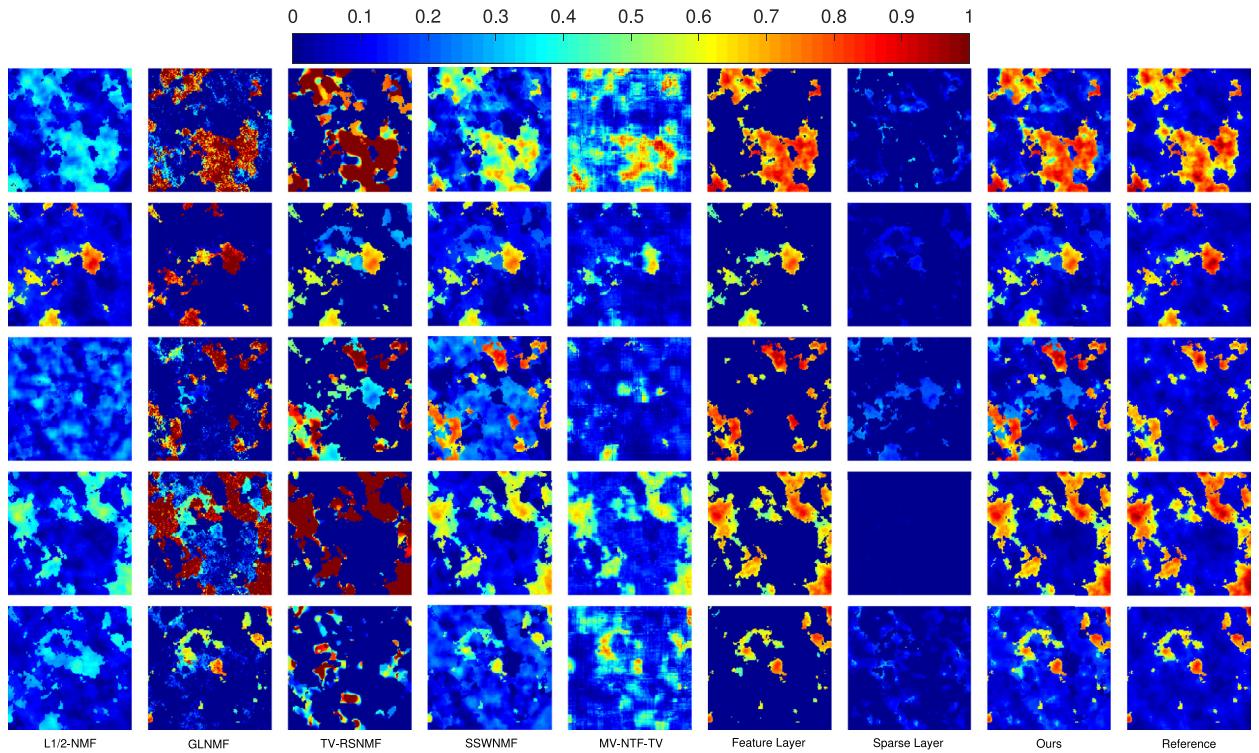


Fig. 6. Estimated abundance maps by different unmixing algorithms on DC1.

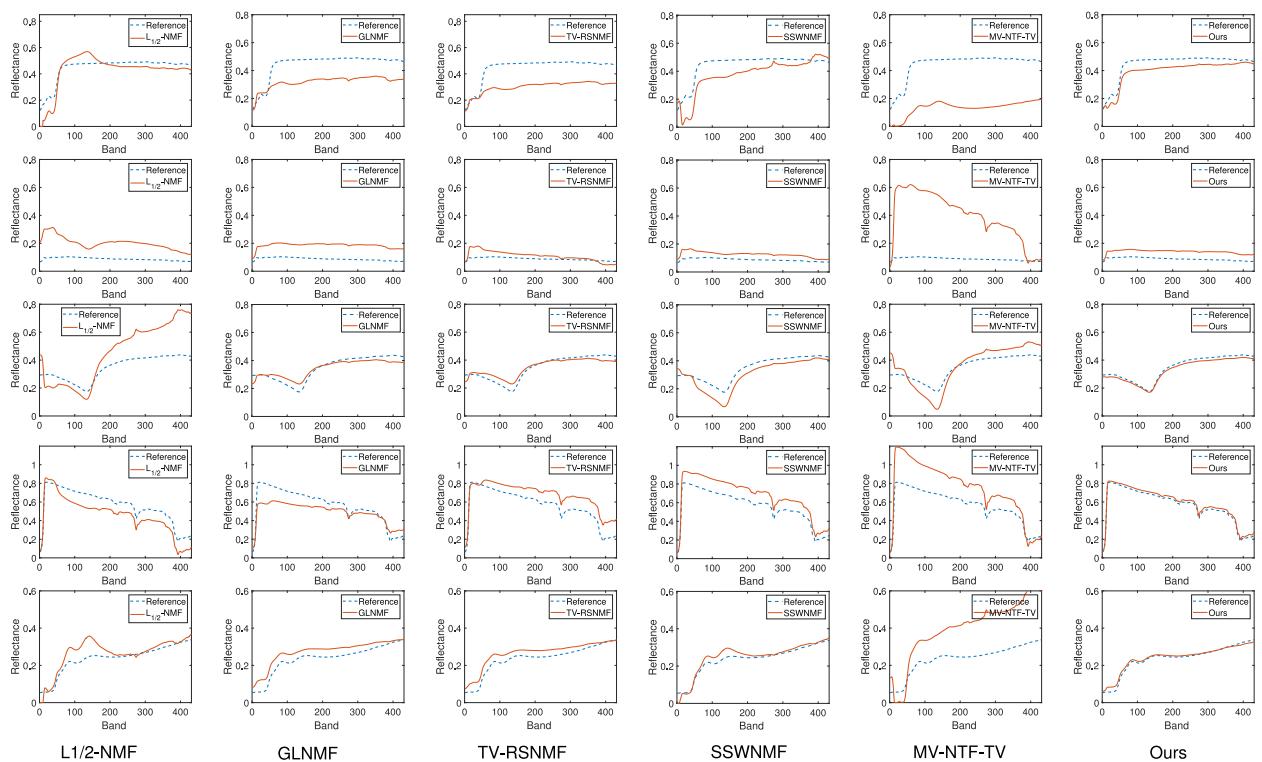
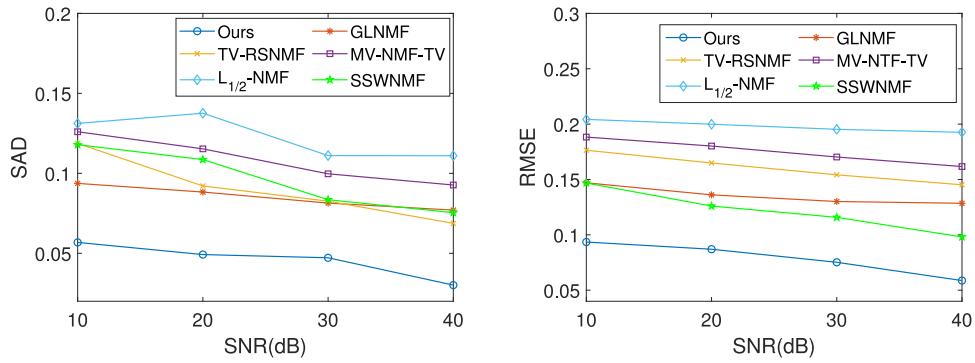


Fig. 7. True and estimated spectral curves by different unmixing algorithms on DC1.

Table 1

The SAD and RMSE of different algorithms on DC1.

Algorithm		$L_{1/2}$ -NMF	GLNMF	TV-RSNMF	SSWNMF	MV-NTF-TV	Ours
SAD	End 1	0.0972	0.1064	0.0970	0.1597	0.1816	0.0627
	End 2	0.1465	0.0865	0.0631	0.0214	0.0768	0.0132
	End 3	0.2088	0.0554	0.0850	0.0727	0.0528	0.0473
	End 4	0.1095	0.0891	0.0890	0.1207	0.1235	0.0041
	End 5	0.1003	0.0696	0.0631	0.0870	0.2050	0.0401
	Mean	0.1324	0.0814	0.0782	0.0870	0.0952	0.0335
RMSE	Mean	0.1926	0.1285	0.1450	0.1001	0.1617	0.0587

**Fig. 8.** Values of SAD and RMSE obtained by different unmixing algorithms with different SNRs on DC1.

feature extraction regularization, called SFE-NTF-TV. We show the estimated abundance maps by SFE-NTF, FE-NTF, and SFE-NTF-TV on DC1 in Fig. 5. From Fig. 5, FE-NTF and TV-SFE-TV catch the outline of the abundance maps, but fail in estimating the fractions accurately.

For the endmember 3, the corresponding estimated abundance maps by TV-SFE-NTF are more accurate than FE-NTF. Compared with FE-NTF and TV-SFE-TV, SFE-NTF estimates the abundance maps better. That shows the effectiveness of Ostu's algorithm to explore the information of the feature layer.

4.3. Experiments on DC1

1) Performance of comparing algorithms: We now compare the unmixing performance of the six algorithms on DC1. Table 1 shows the values of SAD and RMSE for quantitative evaluation. Specifically, considering that there are small deviations among the endmembers estimated by the comparing algorithms, we show the values of SAD for each endmember. From Table 1, we observed that our algorithm achieves better SAD and RMSE than other algorithms. Meanwhile, GLNMF, TV-RSNMF and SSWNMF obtain better SAD and RMSE than $L_{1/2}$ -NMF.

To visibly show the unmixing performance, Figs. 6 and 7 present the estimated abundance maps and spectral curves by the comparing algorithms on DC1, respectively. From Fig. 6, all the comparing algorithms recover the abundance maps of endmembers 2 and 4 well. And GLNMF, TV-RSNMF, and our algorithm provide closer abundance maps of endmembers 1 and 3 to the references. However, in the abundance maps of endmembers 1, 3, and 5 estimated by GLNMF, TV-RSNMF, there are some values which are not present in the references.

SSWNMF recovers the most information of the abundance maps, but fails in removing noise in the backgrounds.

Clearly, our algorithm alleviates this problem and provides more accurate abundance maps, which demonstrate the effectiveness of separating each abundance map into the feature layer and the sparse layer. From Fig. 7, the endmember spectral curves estimated by GLNMF, TV-RSNMF, and SSWNMF are closer to the references than those obtained by $L_{1/2}$ -NMF. Comparing with other algorithms, our algorithm provides closer spectral curves to the references for endmembers 1 to 5.

2) Robustness to Noise: We now test the noise sensitivity of the comparing algorithms on DC1 with SNR = 10, 20, 30, and 40 dB. Fig. 8 shows values of SAD and RMSE obtained by the comparing algorithms with different SNRs on DC1. From Fig. 8, with the noise level increases, values of SAD and RMSE computed by these six algorithms become larger. Additionally, our algorithm achieves the best SAD and RMSE comparing with other algorithms under all noise levels.

To further show the noise sensitivity of these six algorithms with different SNRs, we present the estimated abundance maps of endmember 3 in Fig. 9. Similar conclusions are obtained for other endmembers, thus we omit them here for space consideration. According to Fig. 9, $L_{1/2}$ -NMF recover the outline of the abundance maps with different noise levels, but provide noisy backgrounds. The other four algorithms not only catch the outline of abundance maps, but also delineate the high fractional abundance regions under different noise levels. It demonstrates that exploiting spatial information is

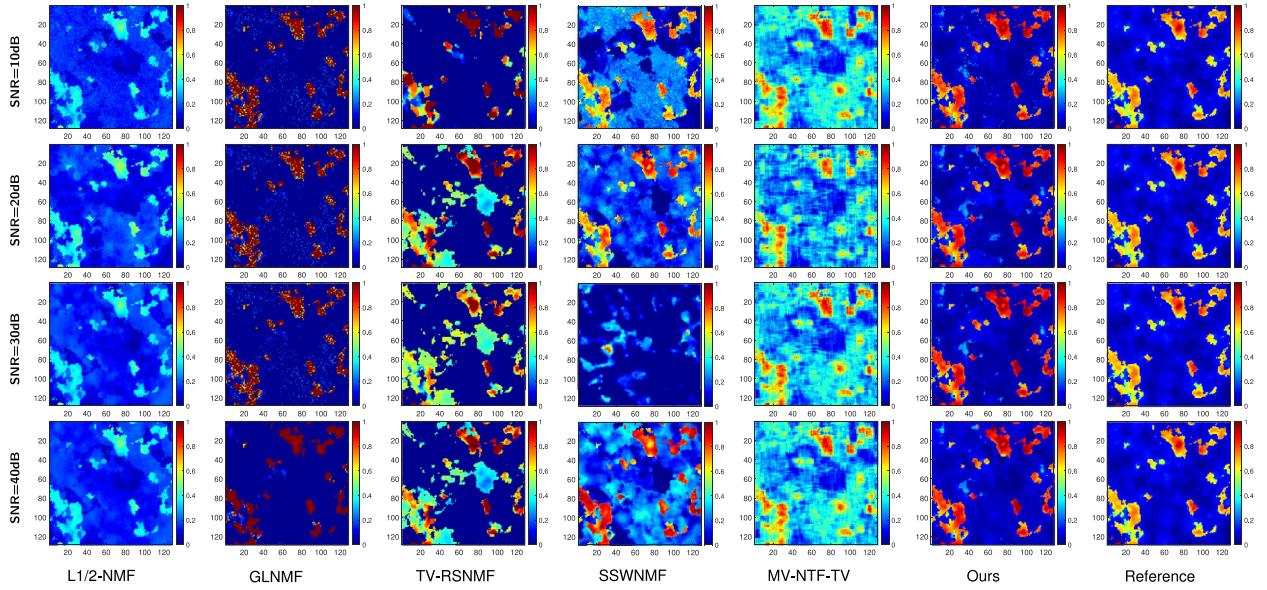


Fig. 9. Estimated abundance maps by different unmixing algorithms with different SNRs on DC1.

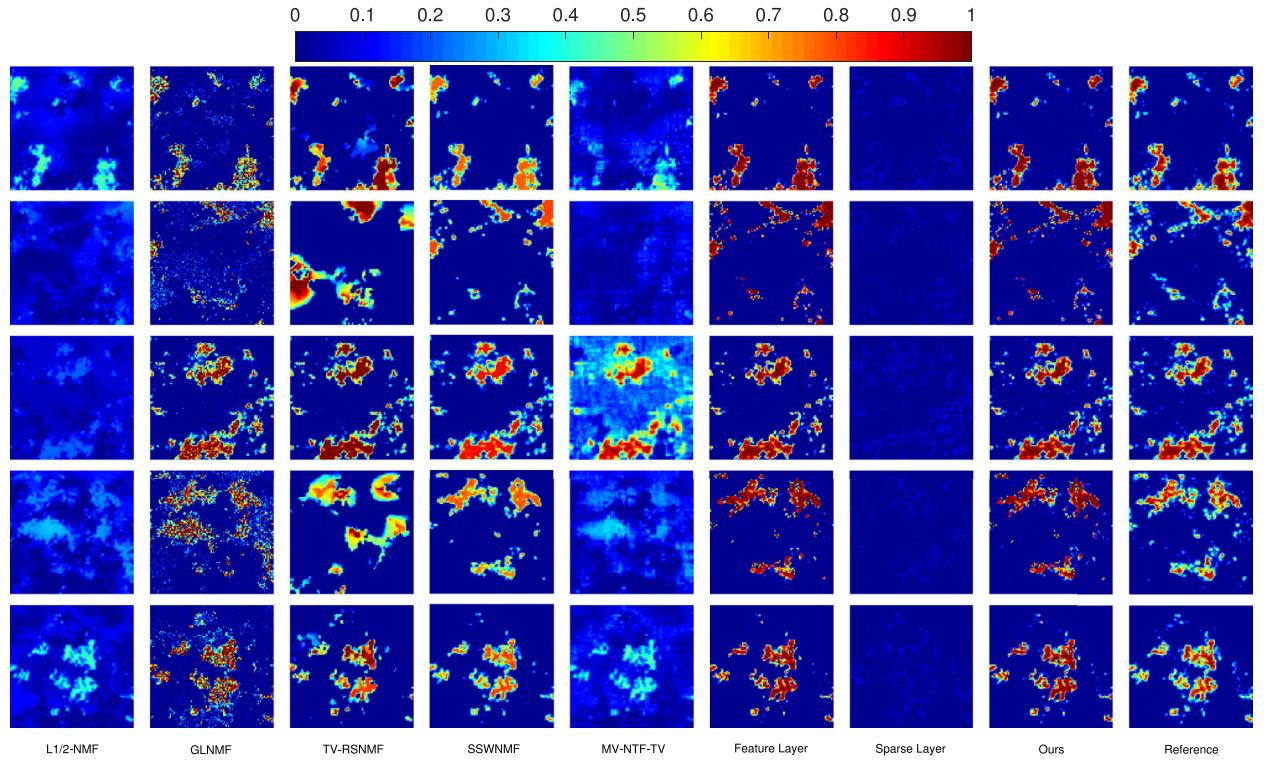


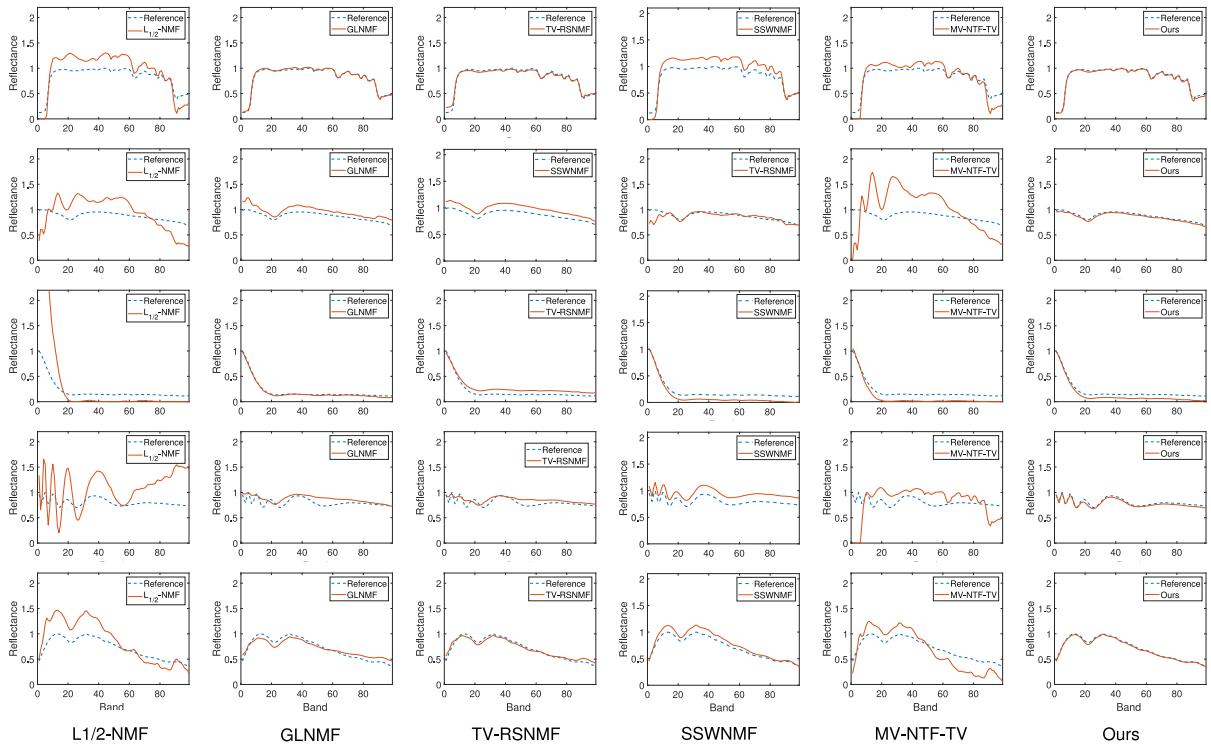
Fig. 10. True and estimated abundance maps by different unmixing algorithms for endmembers 1, 3, 5, 6, and 8 on DC2.

effective for obtaining better unmixing performance. However, GLNMF and SSWNMF suffer from noise and produce some small abundances which are not in the references. TV-RSNMF removes some edge pixels with $\text{SNR} = 40 \text{ dB}$ and generates some abundances in the neighbors of the edge pixels with $\text{SNR} = 10, 20, \text{ and } 30 \text{ dB}$. Similarly, MV-NTF-TV generates some values which are not present in the references. Under different noise levels, our algorithm provides closer abundance maps to the references comparing with other algorithms.

Table 2

Values of SAD and RMSE obtained by different algorithms on DC2.

Algorithm	$L_{1/2}$ -NMF	GLNMF	TV-RSNMF	SSWNMF	MV-NTF-TV	Ours
SAD	End 1	0.1320	0.0148	0.0338	0.0288	0.1257
	End 2	0.2839	0.2425	0.2598	0.3305	0.2563
	End 3	0.2715	0.0342	0.0382	0.0443	0.2304
	End 4	0.1742	0.0191	0.0338	0.0146	0.1905
	End 5	0.4824	0.0455	0.0777	0.0131	0.4476
	End 6	0.2622	0.0680	0.0285	0.0261	0.2100
	End 7	0.2220	0.0449	0.0799	0.0085	0.1312
	End 8	0.1830	0.0872	0.0347	0.0118	0.1623
	End 9	0.1541	0.0431	0.0398	0.0159	0.1569
	Mean	0.2405	0.0666	0.0695	0.0547	0.1771
RMSE	Mean	0.2412	0.1487	0.1374	0.0842	0.1689
	Mean					0.0753

**Fig. 11.** True and estimated spectral curves by different unmixing algorithms for endmembers 1, 3, 5, 6, and 8 on DC2.

4.4. Experiments on DC2

To show the unmixing performance under different types of statistical distributions, we apply the comparing algorithms on DC2 and show the quantitative results in [Table 2](#). From [Table 2](#), GLNMF, TV-RSNMF, SSWNMF and our algorithm achieve lower SADs and RMSEs than $L_{1/2}$ -NMF. Notice that, our algorithm also obtains better RMSE than other comparing algorithms, which demonstrates the advantage of our algorithm in estimating spectral curves and abundance maps.

Furthermore, [Figs. 10](#) and [11](#) show the estimated abundance maps and spectral curves of endmembers 1, 3, 5, 6, and 8 by six algorithms for visual comparison. The spectral curves and the abundance maps of other endmembers demonstrate similar performances, we omit them here for space considerations. By [Fig. 10](#), $L_{1/2}$ -NMF catches the outline of the abundance maps for endmembers 1, 5, and 8, but fail in the abundance maps of endmembers 3 and 6. GLNMF and TV-RSNMF estimate abundance maps of endmembers 1, 5, and 8 better than $L_{1/2}$ -NMF and recover the outline of the abundance maps of endmembers 3 and 6. However, the abundance maps estimated by GLNMF exist noise. TV-RSNMF smooths the edge pixels, which causes that the neighbors of the edge pixels are assigned some values. SSWNMF estimates the abundance maps better than GLNMF and TV-RSNMF. Comparing with the above five algorithms, our algorithm achieves closer abundance maps of all endmembers to references. According to [Fig. 11](#), GLNMF, TV-RSNMF, SSWNMF, and our algorithm obtain comparable spectral curves for all endmembers. Specifically, the spectral curves of endmembers 5 and 8 estimated by GLNMF,

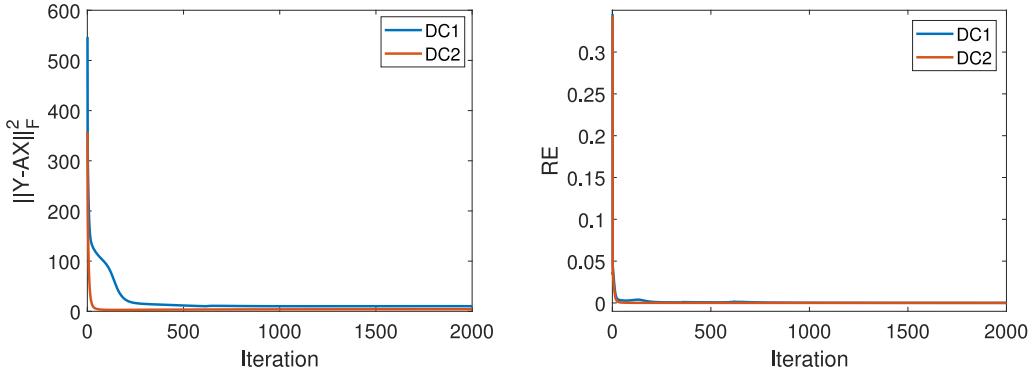


Fig. 12. Convergence curves of our algorithm on DC1 and DC2.

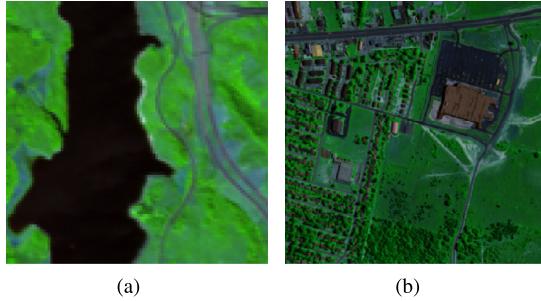


Fig. 13. The false-color images for Jasper Ridge data (a) and Urban data (b).

TV-RSNMF, SSWNMF, and our algorithm are similar. Generally, GLNMF, TV-RSNMF, and SSWNMF estimate spectral curves more accurate than $L_{1/2}$ -NMF. It demonstrates that exploiting spatial information is essential for estimating spectral curves. In conclusion, our algorithm provides more accurate estimations than other algorithms.

4.5. Convergence analysis

In [47,48], it has been pointed out that the multiplicative iterative algorithm is hard to converge to a stationary point. Thus, this experiment is to numerically analyze the convergence of SFE-NTF. In Fig. 12, we plot the curves of the values of $\|\mathbf{Y} - \mathbf{AX}\|_F^2$ and the relative error (RE) of \mathbf{E} versus the iteration number on DC1 and DC2. The RE of \mathbf{E} is defined as follows:

$$\text{RE} = \frac{\|\mathbf{E}^{(t+1)} - \mathbf{E}^{(t)}\|_F}{\|\mathbf{E}^{(t)}\|_F}.$$

From Fig. 12, the convergence curves of SFE-NTF tend to be stable on two data cubes.

5. Real-World data experiments

In this section, two real-world HSI data sets are adopted to show the unmixing effectiveness of our algorithm in practical cases. These two data sets are Jasper Ridge data set and Urban data set, which are widely used in the field of HSI processing [49,50]. We adopt the same initial strategy to the synthetic experiments. In terms of the parameters in SFE-NTF, we fix $\delta = 5$, $\mu = 1$, and λ , μ are selected from $\{0.3, 0.5, 1, 3, 5\}$ in the following experiments.

5.1. Jasper ridge data set

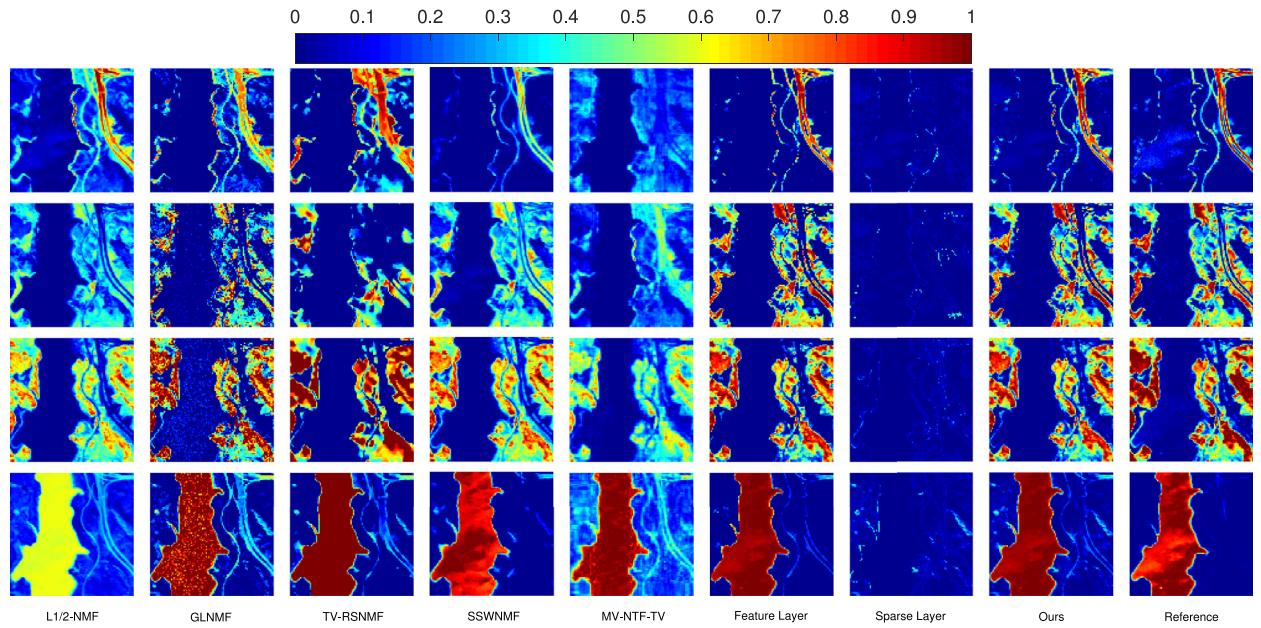
Jasper Ridge data set is from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Jasper Ridge in central California, USA [51]. The original Jasper Ridge data set contains 512×614 pixels with 224 bands, whose reflectance values distribute uniformly ranging from 0.38 to $2.5 \mu\text{m}$. Considering the effect of water-absorption and noise, we delete bands 1 – 3, 108 – 112, 154 – 166, and 220 – 224 and leave 198 channels in this experiment. Similarly as in [17,21], we adopt a 100×100 -pixel sub-image of Jasper Ridge data set, which contains four endmembers: road, soil, water, and tree. Fig. 13(a) shows the false-color image of the Jasper Ridge data.

Quantitative results on the Jasper Ridge data set are shown in Table 3. From Table 3, GLNMF and TV-RSNMF obtain lower SAD for endmember tree. And our algorithm's SADs are lower than others for endmembers water, soil, and road. In addition, our algorithm achieves the best mean SAD on the Jasper Ridge data set.

Table 3

Values of SAD obtained by different algorithms on Real-World Data.

Jasper Ridge data						
Algorithm	$L_{1/2}$ -NMF	GLNMF	TV-RSNMF	SSWNMF	MV-NTF-TV	Ours
Tree	0.3012	0.0239	0.0277	0.1525	0.2261	0.0412
Water	0.2132	0.1007	0.0830	0.1072	0.0456	0.0253
Soil	0.2616	0.1181	0.1365	0.1600	0.1732	0.0834
Road	0.0611	0.1151	0.1455	0.1383	0.1153	0.0845
Mean	0.1930	0.0894	0.0979	0.1395	0.1375	0.0560
Urban data						
Algorithm	$L_{1/2}$ -NMF	GLNMF	TV-RSNMF	SSWNMF	MV-NTF-TV	Ours
Asphalt	0.1134	0.1506	0.1240	0.1276	0.0456	0.1171
Grass	0.1916	0.0791	0.0960	0.1026	0.1153	0.0462
Tree	0.4196	0.1215	0.1275	0.0531	0.1732	0.0943
Roof	0.2280	0.1118	0.1213	0.1276	0.2261	0.0437
Mean	0.2381	0.1157	0.1172	0.1337	0.2015	0.0753

**Fig. 14.** True and estimated abundance maps by different unmixing algorithms for road, soil, tree, and water on Jasper Ridge data.

We show the estimated abundances of the comparing algorithms on the Jasper Ridge data set in Fig. 14. Generally, the pure region appears red. As Fig. 14 shown, all six algorithms recover the most information of the abundance maps. However, $L_{1/2}$ -NMF fails in catching the accurate fractional abundances. The abundance maps obtained by GLNMF are noisy and exist some regions which are not in the references. TV-RSNMF provides smoother abundance maps comparing with other algorithms, but misses many details in the abundance maps. SSWNMF provides more accurate abundance maps than GLNMF and TV-RSNMF. Specifically, for the abundance map of endmember road, our algorithms provides a closer estimation to the reference. In conclusion, our algorithm achieves closer estimations to references for all abundance maps. In addition, Fig. 15 shows the spectral curves estimated by the six algorithms on the Jasper Ridge data set. According to Fig. 15, the spectral curves for endmember water estimated by TV-RSNMF, MV-NTF-TV, SSWNMF, and our algorithm are similar and closer to the reference. For the spectral curves of endmembers road and soil, our results are closer to the references. Conclusively, our algorithm produces more accurate spectral curves comparing with other algorithms.

5.2. Urban data set

The Urban data set is generated by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) in an urban area [52]. This data set contains 307×307 pixels with 210 bands, whose reflectance values are in the range from 0.4 to $2.5 \mu\text{m}$. Removing the bands contaminated by dense water vapor and noise, we remain 162 bands. In this experiment, we assume that 4 endmembers: asphalt, grass, roof, tree participate in the mixing process. Fig. 13(b) shows the false-color image of the Urban data set.

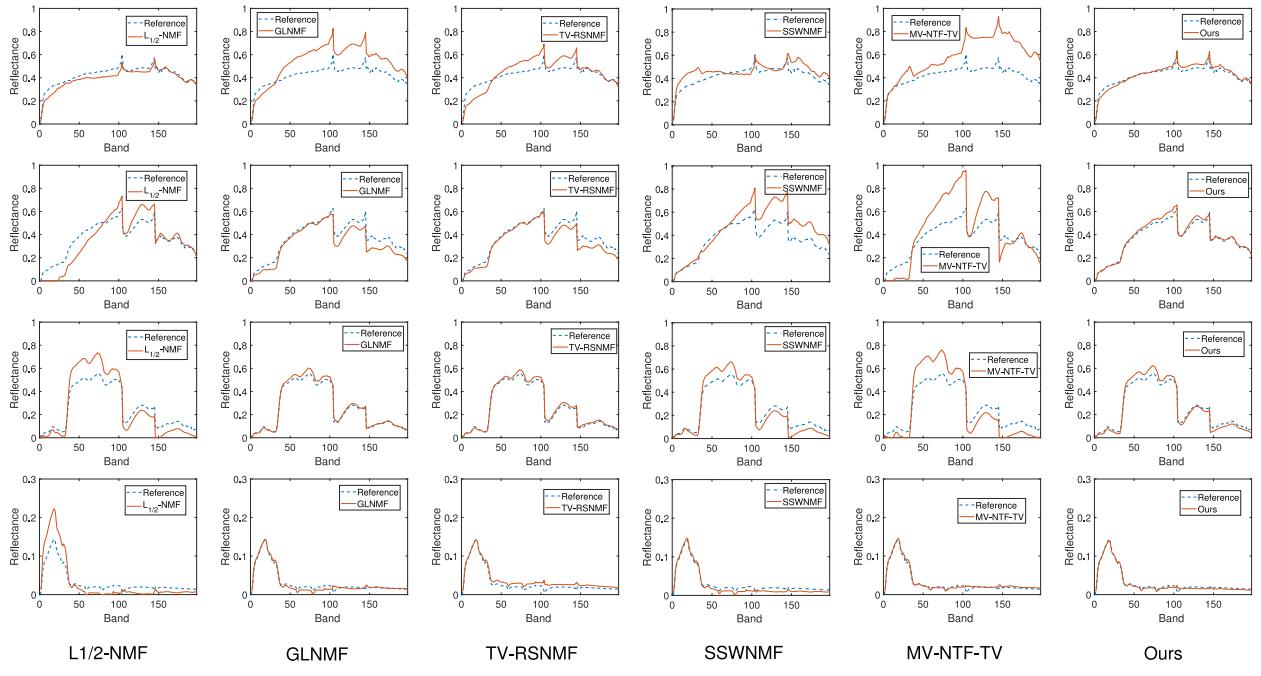


Fig. 15. True and estimated spectral curves by different unmixing algorithms for road, soil, tree, and water on Jasper Ridge data.

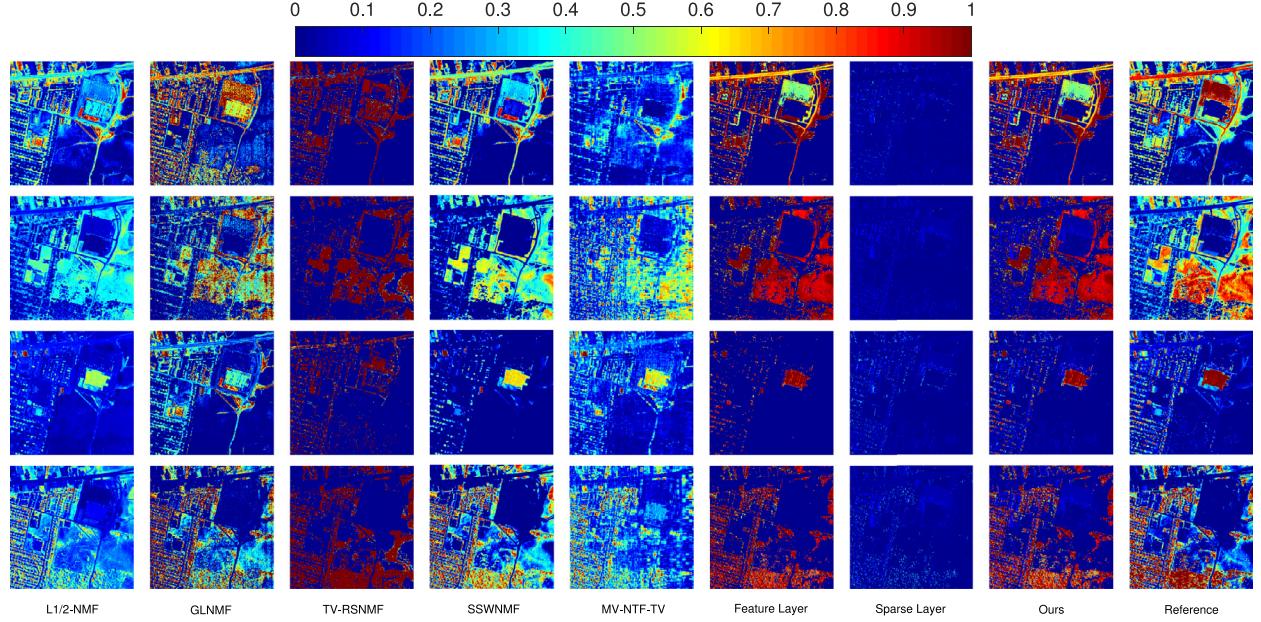


Fig. 16. True and estimated abundance maps by different unmixing algorithms for asphalt, grass, roof, and tree on Urban data.

Table 3 shows the SADs obtained by comparing six algorithms on the Urban data set. From **Table 3**, MV-NTF-TV obtains a lower SAD for endmember asphalt, and our algorithm achieves lower SADs for endmembers grass, tree, and roof.

Furthermore, we also present the abundance maps and the spectral curves recovered by the six algorithms in Figs. 16 and 17. According to Fig. 16, the six algorithms recover the most information of the abundance maps. Specifically, $L_{1/2}$ -NMF, GLNMF, TV-RSNMF, and SSWNMF separate the abundance maps of endmembers asphalt and roof inadequately. For the other abundance maps, $L_{1/2}$ -NMF provides the low fractional abundances. TV-RSNMF obtains high fractional abundances, but separates endmembers grass and tree inadequately. SSWNMF catches the most information of the abundance maps, but the most estimated fractions by SSWNMF are lower than the references. As the estimated spectral curves showed in Fig. 17, the spectral curves of endmember asphalt estimated by MV-NTF-TV is closer to the reference. And our algorithm

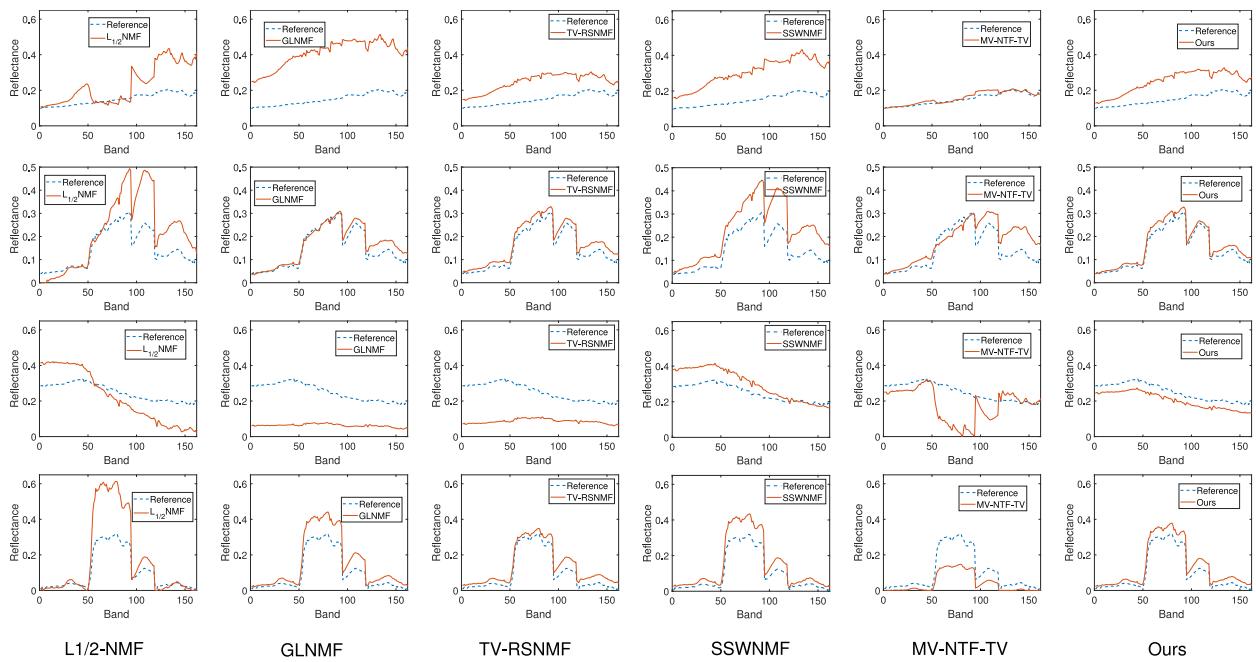


Fig. 17. True and estimated spectral curves by different unmixing algorithms for asphalt, grass, roof, and tree on Urban data.

provides more accurate spectral curves for other endmembers comparing with the other algorithms. It is consistent with the conclusion in the Table 3. Compared with these five algorithms, our algorithm gives closer estimations to the references in most cases.

6. Conclusions

In this paper, we propose a new unmixing model named spatial feature extraction non-negative tensor factorization (SFE-NTF) for hyperspectral unmixing. In SFE-NTF, we separate each abundance map into the feature layer and the sparse layer to describe the two different parts of each abundance map. The feature layer is the part belonging to the true abundance map, and the sparse layer is the part containing some small abundances, which are caused by factors like noise. We introduce a spatial feature extraction regularization into the feature layer to catch the most information of the abundance maps. And we exploit the sparse layer with the weighted ℓ_1 norm. Experimental results on both synthetic data and real-world data demonstrate the unmixing effectiveness of the proposed SFE-NTF.

Acknowledgments

We would like to thank the authors of $L_{1/2}$ -NMF, GLNMF, TV-RSNMF, SSWNMF, and MV-NTF-TV for sharing their codes. This research is supported in part by NSFC (Grant No. 12171072, 61772003), in part by Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216), in part by National Key Research and Development Program of China (Grant No. 2020YFA0714001), and in part by the Fundamental Research Funds for the Central Universities (Grant No. ZYJX2019J093).

Last but not least, the authors would like to thank the editors and the two anonymous reviewers for their constructive, detailed, and helpful advice regarding this paper.

References

- [1] P.S. Thenkabail, R.B. Smith, E.D. Pauw, Hyperspectral vegetation indices and their relationships with agricultural crop characteristics, *Remote. Sens. Environ.* 71 (2) (2000) 158–182.
- [2] Y. Liu, Y.R. Chen, C.Y. Wang, D.E. Chan, M.S. Kim, Development of hyperspectral imaging technique for the detection of chilling injury in cucumbers, *Proc. SPIE* 22 (1) (2004) 101–114.
- [3] M. Kamruzzaman, G. Elmairy, D.W. Sun, P. Allen, Application of nir hyperspectral imaging for discrimination of lamb muscles, *J. Food. Eng.* 104 (3) (2011) 332–340.
- [4] L.J. Deng, M.Y. Feng, X.C. Tai, The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-laplacian prior, *Inf. Fusion* 52 (2019) 76–89.
- [5] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P.D. Gader, J. Chanussot, Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches, *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.* 5 (2) (2012) 354–379.
- [6] N. Keshava, J.F. Mustard, Spectral unmixing, *IEEE Signal Process. Mag* 19 (1) (2002) 44–57.

- [7] W.B. Joseph, A.K. Fred, O.G. Robert, Mineral mapping with hyperspectral digital imagery collection experiment (hydice) sensor data at cuprite, nevada, usa., *Int. J. Remote Sens.* 18 (74) (1997) 1553–1570.
- [8] M.E. Winter, N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, *Proc. SPIE* 3753 (1999) 266–275.
- [9] J.M.P. Nascimento, J. Dias, Vertex component analysis: a fast algorithm to unmix hyperspectral data, *IEEE Trans. Geosci. Remote Sens.* 43 (4) (2005) 898–910.
- [10] L. Zhuang, C.H. Lin, M.A.T. Figueiredo, J.M. Bioucas-Dias, Regularization parameter selection in minimum volume hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 57 (12) (2019) 9858–9877.
- [11] C.H. Lin, J.M. Bioucas-Dias, Nonnegative blind source separation for ill-conditioned mixtures via john ellipsoid, *IEEE Trans. Neural Netw. Learn. Syst.* 32 (5) (2021) 2209–2223.
- [12] S. Jia, Y.T. Qian, Constrained nonnegative matrix factorization for hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 47 (1) (2009) 161–173.
- [13] Z.Y. Yang, G.X. Zhou, S.L. Xie, S.X. Ding, J.M. Yang, J. Zhang, Blind spectral unmixing based on sparse nonnegative matrix factorization, *IEEE Trans. Image Process* 20 (4) (2011) 1112–1125.
- [14] M. Iordache, J.M. Bioucas-Dias, A. Plaza, Sparse unmixing of hyperspectral data, *IEEE Trans. Geosci. Remote Sens.* 49 (6) (2011) 2014–2039.
- [15] J. Huang, T.Z. Huang, L.J. Deng, X.L. Zhao, Joint-sparse-blocks and low-rank representation for hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 57 (4) (2019) 2419–2438.
- [16] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: *Proc. 13th Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 535–541.
- [17] Y.T. Qian, F.C. Xiong, S. Zeng, J. Zhou, Y.Y. Tang, Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery, *IEEE Trans. Geosci. Remote Sens.* 55 (3) (2017) 1776–1792.
- [18] L. Tong, J. Zhou, X. Li, Y. Qian, Y. Gao, Region-based structure preserving nonnegative matrix factorization for hyperspectral unmixing, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 10 (4) (2017) 1575–1588.
- [19] V.P. Pauca, J. Piper, R.J. Plemmons, Nonnegative matrix factorization for spectral data analysis, *Linear Algebra Appl.* 416 (1) (2006) 29–47.
- [20] D.L. Donoho, V. Stodden, When does non-negative matrix factorization give a correct decomposition into parts? *Proc. Adv. Neural Inf. Process. Syst. Conf.* (2003) 1141–1148.
- [21] F.C. Xiong, Y.T. Qian, J. Zhou, Y.Y. Tang, Hyperspectral unmixing via total variation regularized nonnegative tensor factorization, *IEEE Trans. Geosci. Remote Sens.* 57 (4) (2019) 2341–2357.
- [22] W. He, H.Y. Zhang, L.P. Zhang, Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 9 (9) (2016) 4267–4279.
- [23] C.G. Tsinos, A.A. Rontogiannis, K. Berberidis, Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization, *IEEE Trans. Comput. Imag* 3 (2) (2017) 160–174.
- [24] Y.T. Qian, S. Jia, J. Zhou, A. Robles-Kelly, Hyperspectral unmixing via $\ell_{1/2}$ sparsity-constrained nonnegative matrix factorization, *IEEE Trans. Geosci. Remote Sens.* 49 (11) (2011) 4282–4297.
- [25] F.Y. Zhu, Y. Wang, S.M. Xiang, B. Fan, C.H. Pan, Structured sparse method for hyperspectral unmixing, *ISPRS J. Photogramm. Remote Sens.* 88 (2014) 101–118.
- [26] X.Q. Lu, H. Wu, Y. Yuan, P.K. Yan, X.L. Li, Manifold regularized sparse nmf for hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 51 (5) (2013) 2815–2826.
- [27] L. Qi, J. Li, Y. Wang, Y.F. Huang, X.B. Gao, Spectral-spatial-weighted multiview collaborative sparse unmixing for hyperspectral images, *IEEE Trans. Geosci. Remote Sens.* 58 (12) (2020) 8766–8779.
- [28] W. He, H.Y. Zhang, L.P. Zhang, Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 55 (7) (2017) 3909–3921.
- [29] X.L. Zhao, W. Wang, T.Y. Zeng, T.Z. Huang, M.K. Ng, Total variation structured total least squares method for image restoration, *SIAM J Sci Comput.* 35 (6) (2013) B1304–B1320.
- [30] W.H. Wang, Y.T. Qian, Y.Y. Tang, Hypergraph-regularized sparse nmf for hyperspectral unmixing, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 9 (2) (2016) 681–694.
- [31] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, *SIAM Rev.* 51 (3) (2009) 455–500.
- [32] N. Senthilkumaran, S. Vaithogi, Image segmentation by using thresholding techniques for medical images, *Computer Science & Engineering: An International Journal* 6 (2016) 1–13.
- [33] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: algorithms and applications, *IEEE Access* 3 (2015) 490–530.
- [34] E.J. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (12) (2005) 4203–4215.
- [35] E.J. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory* 52 (12) (2006) 5406–5425.
- [36] E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.* 14 (2008) 877–905.
- [37] J.F. Cai, S. Osher, Z. Shen, Convergence of the linearized bregman iteration for ℓ_p -norm minimization, *Math. Comput.* 78 (268) (2009) 2127–2136.
- [38] W. Yin, D.G. S. Osher, J. Darbon, Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing, *SIAM J. Imag. Sci.* 1 (1) (2008) 143–168.
- [39] S.Q. Zhang, G.R. Zhang, F. Li, C.Z. Deng, S.Q. Wang, A. Plaza, J. Li, Spectral-spatial hyperspectral unmixing using nonnegative matrix factorization, *IEEE Trans. Geosci. Remote Sens.* (2021) 1–13, doi:[10.1109/TGRS.2021.3074364](https://doi.org/10.1109/TGRS.2021.3074364).
- [40] J.M. P Nascimento, J.M. Bioucas-Dias, Hyperspectral signal subspace estimation, in: *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 3225–3228.
- [41] L. Dong, Y. Yuan, X.Q. Luxs, Spectral-spatial joint sparse nmf for hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 59 (3) (2021) 2391–2402.
- [42] S.Q. Zhang, J. Li, Z.B. Wu, A. Plaza, Spatial discontinuity-weighted sparse unmixing of hyperspectral images, *IEEE Trans. Geosci. Remote Sens.* 56 (10) (2018) 5767–5779.
- [43] J.J. Wang, T.Z. Huang, J. Huang, H.X. Dou, L.J. Deng, X.L. Zhao, Row-sparsity spectral unmixing via total variation, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 12 (12) (2019) 5009–5022.
- [44] J. Huang, T.Z. Huang, X.L. Zhao, L.J. Deng, Nonlocal tensor-based sparse hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 59 (8) (2021) 6854–6868.
- [45] K. Abercromby, Communication of the NASA JSC Spacecraft Materials Spectral Database, 2006.
- [46] C.J. Lin, On the convergence of multiplicative update algorithms for nonnegative matrix factorization, *IEEE Trans. Neural Netw.* 18 (6) (2007) 1589–1596.
- [47] E. Gonzalez, Y. Zhang, Accelerating the lee-seung algorithm for non-negative matrix factorization, Dept. Comput. Appl. Math., Rice Univ., Houston, TX, USA, Tech. Rep (2005) 1–13.
- [48] F.Y. Zhu, Y. Wang, B. Fan, S.M. Xiang, G.F. Meng, C.H. Pan, Spectral unmixing via data-guided sparsity, *IEEE Trans. Image Processing* 23 (12) (2014) 5412–5427.
- [49] F.Y. Zhu, Y. Wang, B. Fan, G.Y. Meng, C.H. Pan, Effective spectral unmixing via robust representation and learning-based sparsity, arXiv: Proc. Int. Conf. Comput. Vis. Pattern Recognit. (CVPR) (2014).
- [50] AVIRIS, Aviris Jasper Data, 2014. Available-online: (<http://www.escience.cn/system/file?fileId=68574>).
- [51] HYDICE, Hydice Urban Data, 2014. Available-online: (<http://www.escience.cn/system/file?fileId=68574>).