

# Manifold Regularized Sparse NMF for Hyperspectral Unmixing

Xiaoqiang Lu, Hao Wu, Yuan Yuan, *Senior Member, IEEE*, Pingkun Yan, *Senior Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

**Abstract**—Hyperspectral unmixing is one of the most important techniques in analyzing hyperspectral images, which decomposes a mixed pixel into a collection of constituent materials weighted by their proportions. Recently, many sparse *nonnegative matrix factorization* (NMF) algorithms have achieved advanced performance for hyperspectral unmixing because they overcome the difficulty of absence of pure pixels and sufficiently utilize the sparse characteristic of the data. However, most existing sparse NMF algorithms for hyperspectral unmixing only consider the Euclidean structure of the hyperspectral data space. In fact, hyperspectral data are more likely to lie on a low-dimensional submanifold embedded in the high-dimensional ambient space. Thus, it is necessary to consider the intrinsic manifold structure for hyperspectral unmixing. In order to exploit the latent manifold structure of the data during the decomposition, manifold regularization is incorporated into sparsity-constrained NMF for unmixing in this paper. Since the additional manifold regularization term can keep the close link between the original image and the material abundance maps, the proposed approach leads to a more desired unmixing performance. The experimental results on synthetic and real hyperspectral data both illustrate the superiority of the proposed method compared with other state-of-the-art approaches.

**Index Terms**—Hyperspectral unmixing, manifold regularization, mixed pixel, nonnegative matrix factorization (NMF).

## I. INTRODUCTION

WITH the constant development of imaging spectroscopy, hyperspectral images collected by imaging spectrometers have captured increasingly rich spatial, spectral, and radial information, which benefits the theoretical research on hyperspectral data analysis [1]. Hyperspectral data generally contain several hundreds of continuous spectral bands with narrow wavelength intervals. However, there widely exist mixed pixels owing to the limited spatial resolution of the sensors and the variant ground surface [2]. Thus, in order to make full use of the data, hyperspectral unmixing has become an essential procedure, which decomposes a mixed pixel into a collection

of constituent materials (also called endmembers) and their relative proportions (also called abundances) [3], [4].

In the past few decades, a large number of algorithms for hyperspectral unmixing have been designed based on the *linear mixing model* (LMM) [5] with the assumption that distinct endmembers are not interfered with each other [6], [7]. In general, conventional algorithms for hyperspectral unmixing involve two steps: *endmember extraction* and *mixed-pixel decomposition*. The algorithms for identifying endmembers include *pixel purity index* (PPI) [8], N-FINDR [9], *vertex component analysis* (VCA) [10], *independent component analysis* (ICA) [11], etc. PPI creatively addresses spectral unmixing as a convex geometry problem and tries to find the endmembers through projection. N-FINDR aims at looking for the pixels that can determine the largest volume by expanding a simplex inside the given data. In comparison with PPI and N-FINDR, VCA is proven to be a more advanced endmember extraction algorithm based on convex geometry. All these algorithms have been successfully applied into hyperspectral unmixing when assuming the presence of pure pixels in the original image. Nevertheless, this assumption is barely true for the collected hyperspectral images, where the pixels are highly mixed. ICA views the problem of unmixing as a *blind source separation* problem and seeks a transformation under the assumption that the constituent materials are statistically independent. However, this assumption of independence cannot be fulfilled, since it violates the truth that the sum of proportion fractions within each pixel is constant [12]. To improve the performance of ICA, ICA with additional constraints has been proposed in [13]. Under LMM, together with two important constraints, i.e., *abundance nonnegativity constraint* (ANC) and *abundance sum-to-one constraint* (ASC), *fully constrained least squares* (FCLS) for inferring the abundances is proposed in [14].

Since the assumption of pure pixels is not reliable, various algorithms are developed from the perspective of statistics to deal with this issue [15]–[18]. Among them, *nonnegative matrix factorization* (NMF) [15] is fairly attractive due to its obvious advantages. NMF attempts to learn a parts-based representation of the data, which is in accordance with how the brain recognizes objects. Utilizing an iterative algorithm, NMF seeks two nonnegative matrices decomposed from the given data. It is worth mentioning that only additive operation is allowed because using subtraction is in contradiction with the intuitive idea of combining parts to constitute a whole. This property exactly guarantees ANC. In recent years, NMF has been applied into hyperspectral unmixing [19]. However, a large quantity of local minima occurs owing to the nonconvexity of the objective

Manuscript received December 3, 2011; revised May 31, 2012; accepted August 4, 2012. Date of publication October 16, 2012; date of current version April 18, 2013. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707104, by the National Natural Science Foundation of China under Grant 61125106, Grant 61100079, Grant 61172142, Grant 61172143, and Grant 91120302, and by the Postdoctoral Science Foundation of China under Grant Y111971400.

X. Lu, Y. Yuan, P. Yan, and X. Li are with the Center for Optical Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China.

H. Wu is with the Faculty of Mathematics and Computer Science, Hubei University, Wuhan 430062, China.

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

Digital Object Identifier 10.1109/TGRS.2012.2213825

function. Besides, since no further constraints are imposed on it, standard NMF is insufficient to produce a desired result.

For the purpose of improving the performance of NMF, a variety of constraints are taken into account. *Nonnegative sparse coding* [20] incorporates sparse coding as sparsity constraint into NMF, whereas smoothness constraint is considered in [21]. These algorithms, however, are not designed for hyperspectral data analysis and, thus, fail to make full use of the characteristics of hyperspectral data, which compromises the performance when being applied into this domain. From a convex geometric point of view, *minimum-volume-constrained NMF* (MVCNMF) [22] utilizes the minimum volume constraint, which drives the virtual endmembers to enclose the data cloud. As discontinuities can be caused by the removal of noisy bands or the unexpected change of endmember abundance, *piecewise smoothness NMF with sparsity constraint* [23] is proposed to dispose of this problem.

Recently, a novel NMF method named  $L_{1/2}$ -NMF [24] has been introduced to unmix hyperspectral data. In [25], the  $L_{1/2}$  regularizer has been shown to own many promising properties, which can provide the best sparse solution among that of the  $L_q$  ( $1/2 \leq q < 1$ ) regularizers, and the sparsity of solution can hardly vary with respect to  $q$  when  $0 < q \leq 1/2$ . By virtue of these properties,  $L_{1/2}$ -NMF leads to a more satisfactory result than that of other  $L_q$ -NMFs. Furthermore, a close relation between  $L_{1/2}$ -NMF and minimum-volume-based methods is illustrated in the sense of geometry. However, all of the aforementioned studies only consider the Euclidean structure of the data space.

Researches have shown that high-dimensional Euclidean space cannot be uniformly *filled up* by the image data. Thus, these image data can be regarded as sampled data from or near a submanifold of an ambient space [26]. Therefore, it is necessary to consider the intrinsic manifold structure while performing hyperspectral unmixing.

Inspired by recent advantages in manifold learning and sparse NMF, a novel algorithm named *graph-regularized  $L_{1/2}$ -NMF* (GLNMF) is proposed for hyperspectral unmixing. Unlike previous algorithms, this proposed algorithm aims at seizing the internal manifold structure of hyperspectral data. Since most existing algorithms deal with the hyperspectral data from a global perspective, such as minimum-volume-based algorithms, they fail to take the local geometrical structure into consideration. However, this class of information is also important due to the close connection between the original image and abundance maps of different endmembers. In order to utilize the intrinsic structural information, GLNMF incorporates sparsity constraint and graph regularization to detect the geometrical structure of the hyperspectral data.

GLNMF is a two-step approach including *sparse constraint* and *graph regularization*. In the first step, the  $L_{1/2}$  regularizer can be adopted as a good candidate for sparsity constraint since the  $L_{1/2}$  regularizer can enforce the sparsity of material abundances. In the second step, a nearest neighbor graph is built to model the local structural information for hyperspectral unmixing. As a result, the graph regularization term plays a role of smooth operator as well, which attenuates the interference of noise. To evaluate the performance, the method is tested on

synthetic and real hyperspectral data, respectively. The results of those experiments demonstrate its effectiveness.

The rest of this paper is organized as follows: Section II gives a brief introduction of the LMM and NMF algorithm. Section III presents GLNMF for unmixing and derives the update algorithm. The experimental results on synthetic data and real-world hyperspectral data are shown in Section IV. Section V concludes this paper.

## II. RELATED WORKS

Here, two crucial concepts, i.e., LMM and NMF, are introduced for a more in-depth discussion. The former lays a solid foundation for linear spectral unmixing schemes, and the latter opens the door to parts-based representation for hyperspectral data.

### A. LMM

As has been mentioned in Section I, LMM assumes that distinct endmembers are not interfered by each other. That is, in a hyperspectral image, the reflectance of a given pixel at band  $l$  is denoted as follows:

$$x_l = \sum_{p=1}^P a_{lp}s_p + e_l \quad (1)$$

where  $a_{lp}$  stands for the spectral radiance of endmember  $p$  at wavelength  $\lambda_l$ ,  $s_p$  denotes the corresponding abundance of endmember  $p$  in the given pixel,  $P$  represents the number of endmembers, and  $e_l$  is the noise. The aforementioned two constraints, i.e., ANC and ASC, can be explicitly given by

$$s_p \geq 0, \quad p = 1, 2, \dots, P \quad (2)$$

$$\sum_{p=1}^P s_p = 1. \quad (3)$$

Adopting matrix notation, a hyperspectral image with  $L$  bands can be described by a matrix  $\mathbf{X} \in \mathbb{R}^{L \times N}$ , where  $N$  is the number of pixels. Hence, LMM can be rewritten as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \quad (4)$$

where  $\mathbf{A} \in \mathbb{R}^{L \times P}$  is an endmember signature matrix,  $\mathbf{S} \in \mathbb{R}^{P \times N}$  denotes the relative endmember abundances, and  $\mathbf{E}$  represents the additional noise. Obviously, each column of  $\mathbf{A}$  corresponds to a type of endmember signature. In practice, the number of endmembers is often much smaller than the number of bands, i.e.,  $P \ll L$ . Consequently, the task of unmixing is to factorize a high-dimensional matrix into two low-dimensional matrices, subjected to ANC and ASC.

### B. NMF

Since NMF was put forward, it has triggered an increasing amount of interest because the method turns out to be applicable to a series of fields.

Using the preceding notations, NMF aims at finding two non-negative matrices  $\mathbf{A} \in \mathbb{R}^{L \times P}$  and  $\mathbf{S} \in \mathbb{R}^{P \times N}$  to approximately represent the given nonnegative matrix  $\mathbf{X} \in \mathbb{R}^{L \times N}$ , which can be written as

$$\mathbf{X} \approx \mathbf{AS} \quad (5)$$

where  $P < \min(L, N)$ . Obviously, this factorization leads to a compressed form of original  $\mathbf{X}$ .

In order to measure the quality of the approximation, a commonly used cost function is the square of the Euclidean distance between  $\mathbf{X}$  and  $\mathbf{AS}$ , i.e.,

$$f(\mathbf{A}, \mathbf{S}) = \|\mathbf{X} - \mathbf{AS}\|_F^2 = \sum_{i,j=1}^{L,N} (\mathbf{X}_{ij} - (\mathbf{AS})_{ij})^2 \quad (6)$$

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

To minimize the cost function  $f$  in (6), an iterative algorithm is presented in [27] as follows:

$$\mathbf{A} \leftarrow \mathbf{A} * \mathbf{XS}^T / \mathbf{ASS}^T \quad (7)$$

$$\mathbf{S} \leftarrow \mathbf{S} * \mathbf{A}^T \mathbf{X} / \mathbf{A}^T \mathbf{AS} \quad (8)$$

where  $(\cdot)^T$  denotes the transpose of the matrix, and  $*$  and  $/$  represent the elementwise multiplication and division, respectively. The convergence of the algorithm has been also proven in [27]. Furthermore, once the initial  $\mathbf{A}$  and  $\mathbf{S}$  are nonnegative, the nonnegativity of the two matrices will keep the same under the rules represented by (7) and (8).

Unfortunately, owing to the nonconvexity of the cost function with respect to  $\mathbf{A}$  and  $\mathbf{S}$  together, there exists a large quantity of local minima. This can be confirmed by the obvious truth that the equality  $\mathbf{AS} = (\mathbf{AD})(\mathbf{D}^{-1}\mathbf{S})$  holds for any invertible matrix  $\mathbf{D}$ . In order to reduce the feasible solution set, sparsity constraints are introduced into standard NMF [20], [21]. A more recent approach named  $L_{1/2}$ -NMF has been proposed [24]. The objective function of  $L_{1/2}$ -NMF is

$$f(\mathbf{A}, \mathbf{S}) = \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda \|\mathbf{S}\|_{\frac{1}{2}} \quad (9)$$

where  $\lambda$  weights the contribution of  $\|\mathbf{S}\|_{(1/2)}$ , and  $\|\mathbf{S}\|_{(1/2)}$  is defined as

$$\|\mathbf{S}\|_{\frac{1}{2}} = \sum_{p=1}^P \sum_{n=1}^N \mathbf{S}_{pn}^{1/2}. \quad (10)$$

However, the estimates of  $L_{1/2}$ -NMF often vary a lot when selecting different initial points. As mentioned in [24],  $L_{1/2}$ -NMF is not robust enough to noise. In this case, it is necessary to exploit the structure of data to stabilize the sparse decomposition.

Overall, when applied into hyperspectral unmixing, the aforementioned studies only consider the sparse distribution of endmembers but neglect the intrinsic geometrical structure of the data, which compromises the performance.

### III. SPARSITY-CONSTRAINED NMF WITH GRAPH REGULARIZATION

Here, the proposed GLNMF can preserve the intrinsic geometrical characteristic of hyperspectral data during the unmixing process. Implementation issues and computational complexity of the algorithm are also discussed.

#### A. Graph-Regularized $L_{1/2}$ -NMF

As mentioned earlier, sparsity-constrained NMF only considers the sparse characteristic and the Euclidean structure of hyperspectral data. However, the intrinsic relation between the data may not be accurately reflected by the Euclidean structure in the high-dimensional data space. Recently, researches have shown that high-dimensional Euclidean space cannot be uniformly *filled up* by the image data. Thus, these image data vary smoothly along the geodesics of the data manifold [26]. Therefore, in order to render better performance, it is necessary to consider the intrinsic manifold structure for hyperspectral unmixing. For convenience, a hyperspectral image with  $L$  bands and  $N$  pixels can be denoted as  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{L \times N}$ . Then, each column of  $\mathbf{X}$  represents the reflectance of an individual pixel. It is known that a real-world hyperspectral image usually covers several hundreds of spectral bands. This implies that each  $\mathbf{x}_n$  can be viewed as a point in  $L$ -dimension space. Recall that the target of unmixing is to seek the underlying endmembers and their relative proportions. Keeping the symbols consistent,  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_P] \in \mathbb{R}^{L \times P}$  and  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \in \mathbb{R}^{P \times N}$  denote the endmember signature matrix and the relative abundances, respectively. Based on the LMM, each column of  $\mathbf{X}$  can be approximately described by

$$\mathbf{x}_n \approx \mathbf{A}\mathbf{s}_n \quad (11)$$

where  $\mathbf{x}_n \in \mathbb{R}^L$ , and  $\mathbf{s}_n \in \mathbb{R}^P$ . When written in the form denoted by (11), it can be regarded that  $\mathbf{s}_n$  is the representation of  $\mathbf{x}_n$  in the new  $P$ -dimension space. In this sense, the column vectors of the matrix  $\mathbf{A}$  make up a set of basic vectors of the new space. Therefore, it establishes a close link between  $\mathbf{x}_n$  and  $\mathbf{s}_n$ , i.e., a link between the original image and abundance maps of different endmembers. One may naturally hope that if the spectral signatures of two pixels are similar, *good* unmixing can preserve this similarity. Hence, an assumption can be generated that if a given  $L$ -dimension data point  $\mathbf{x}_i$  is close to  $\mathbf{x}_j$ , then the corresponding abundance  $\mathbf{s}_i$  is also close to  $\mathbf{s}_j$ . This assumption is referred to as manifold assumption, which can be applied in several image processing fields such as hyperspectral image classification [28], hyperspectral imagery representation [29], and manifold denoising [30]. Inspired by recent works [26], [31], a graph regularizer is incorporated into the sparsity-constrained NMF so as to keep this local invariance assumption in this paper.

Let  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{L \times N}$  be the given data. Then, each  $\{\mathbf{x}_n\}_{n=1}^N$  can represent a data point in  $L$ -dimension space.

TABLE I  
FLOATING-POINT CALCULATION TIMES FOR EACH ITERATION IN GLNMF

	Update <b>A</b>	Update <b>S</b>	Total
addition	$LNP + (L + N)P^2$	$LNP + (L + N)P^2 + (k + 4)PN$	$2LNP + 2(L + N)P^2 + (k + 4)PN$
multiplication	$LNP + (L + N)P^2 + LP$	$LNP + (L + N)P^2 + (k + 2)PN$	$2LNP + 2(L + N)P^2 + (L + 2N + kN)P$
division	$LP$	$NP$	$(L + N)P$

A nearest neighbor graph can be constructed with the data points as its vertices. Denote the weight matrix of the graph as  $\mathbf{W}$ . If  $\mathbf{x}_i$  is one of the  $k$ -nearest neighbors of  $\mathbf{x}_j$ , then the weight is assigned as

$$\mathbf{W}_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}} \quad (12)$$

which is known as the heat kernel [26]. It is apparent that when  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are close,  $\mathbf{W}_{ij}$  is relatively big. According to the above analysis, once  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are close, their new representations  $\mathbf{s}_i$  and  $\mathbf{s}_j$  in the new space should be close too. To achieve so, consider minimizing the following function:

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^N \|\mathbf{s}_i - \mathbf{s}_j\|^2 \mathbf{W}_{ij} &= \sum_{i=1}^N \mathbf{s}_i^T \mathbf{s}_i \mathbf{D}_{ii} - \sum_{i,j=1}^N \mathbf{s}_i^T \mathbf{s}_j \mathbf{W}_{ij} \\ &= \text{Tr}(\mathbf{S} \mathbf{D} \mathbf{S}^T) - \text{Tr}(\mathbf{S} \mathbf{W} \mathbf{S}^T) \\ &= \text{Tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) \end{aligned}$$

where  $\text{Tr}(\cdot)$  represents the trace of a matrix,  $\mathbf{D}$  is a diagonal matrix in which  $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ , and  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ . It is obvious that  $\mathbf{L}$  is symmetric. Incorporating this graph regularization term into the  $L_{1/2}$ -NMF leads to the cost function of *graph-regularized  $L_{1/2}$ -NMF* (GLNMF) as follows:

$$f(\mathbf{A}, \mathbf{S}) = \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda \|\mathbf{S}\|_{\frac{1}{2}} + \frac{\mu}{2} \text{Tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) \quad (13)$$

where  $\mu \geq 0$  is the regularization parameter.

### B. Updating Rules

Similar to the NMF case, the cost function  $f$  in (13) is not convex with respect to  $\mathbf{A}$  and  $\mathbf{S}$  together. Thus, finding the global minima is not realistic. Here, a multiplicative iterative algorithm is derived to cope with this issue.

It is worth noting that minimizing the function  $f$  in (13) is subjected to  $\mathbf{A}_{lp} \geq 0$  and  $\mathbf{S}_{pn} \geq 0$ . Let  $\Psi_{lp} \geq 0$  and  $\Phi_{pn} \geq 0$  be the corresponding Lagrange multipliers. Consider the Lagrange  $\mathcal{L}$  as

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda \|\mathbf{S}\|_{\frac{1}{2}} + \frac{\mu}{2} \text{Tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) \\ &\quad + \text{Tr}(\Psi \mathbf{A}^T) + \text{Tr}(\Phi \mathbf{S}^T) \\ &= \frac{1}{2} \text{Tr}(\mathbf{X} \mathbf{X}^T) - \text{Tr}(\mathbf{X} \mathbf{S}^T \mathbf{A}^T) + \frac{1}{2} \text{Tr}(\mathbf{A} \mathbf{S} \mathbf{S}^T \mathbf{A}^T) \\ &\quad + \lambda \|\mathbf{S}\|_{\frac{1}{2}} + \frac{\mu}{2} \text{Tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) + \text{Tr}(\Psi \mathbf{A}^T) + \text{Tr}(\Phi \mathbf{S}^T) \end{aligned}$$

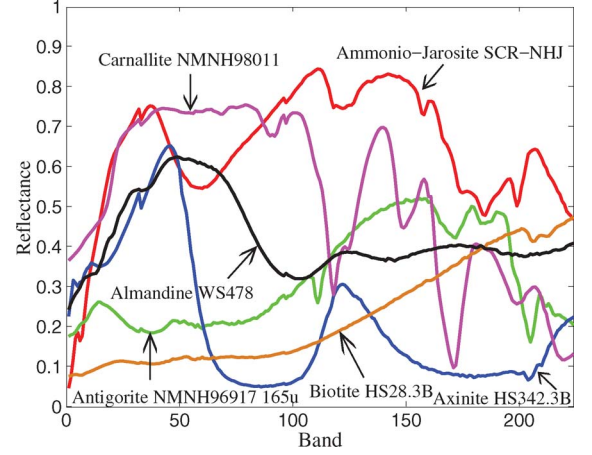


Fig. 1. Example endmember spectra used in synthetic data.

where the second equality utilizes the definition of Frobenius norm  $\|\mathbf{A}\|_F = (\text{Tr}(\mathbf{A}^T \mathbf{A}))^{1/2}$  and two properties of trace  $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$ ,  $\text{Tr}(\mathbf{A}) = \text{Tr}(\mathbf{A}^T)$ . Then, taking the partial derivative with respect to  $\mathbf{A}$  and  $\mathbf{S}$  on both sides yields

$$\frac{\partial \mathcal{L}}{\partial \mathbf{A}} = -\mathbf{X} \mathbf{S}^T + \mathbf{A} \mathbf{S} \mathbf{S}^T + \Psi$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{S}} = -\mathbf{A}^T \mathbf{X} + \mathbf{A}^T \mathbf{A} \mathbf{S} + \frac{1}{2} \lambda \mathbf{S}^{-\frac{1}{2}} + \mu \mathbf{S} \mathbf{L} + \Phi.$$

According to the *Karush-Kuhn-Tucker* conditions  $\Psi_{lp} \mathbf{A}_{lp} = 0$  and  $\Phi_{pn} \mathbf{S}_{pn} = 0$ , it follows that

$$(-\mathbf{X} \mathbf{S}^T + \mathbf{A} \mathbf{S} \mathbf{S}^T) \cdot \mathbf{A} = 0$$

$$\left( -\mathbf{A}^T \mathbf{X} + \mathbf{A}^T \mathbf{A} \mathbf{S} + \frac{1}{2} \lambda \mathbf{S}^{-\frac{1}{2}} + \mu \mathbf{S} \mathbf{L} \right) \cdot \mathbf{S} = 0.$$

By applying transposition and division, together with using  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ , the updating rules can be acquired as

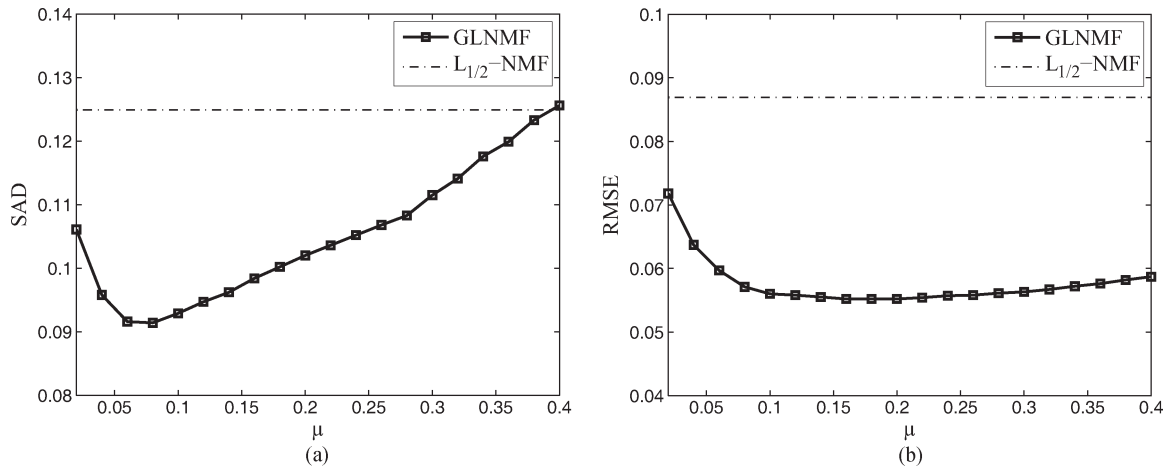
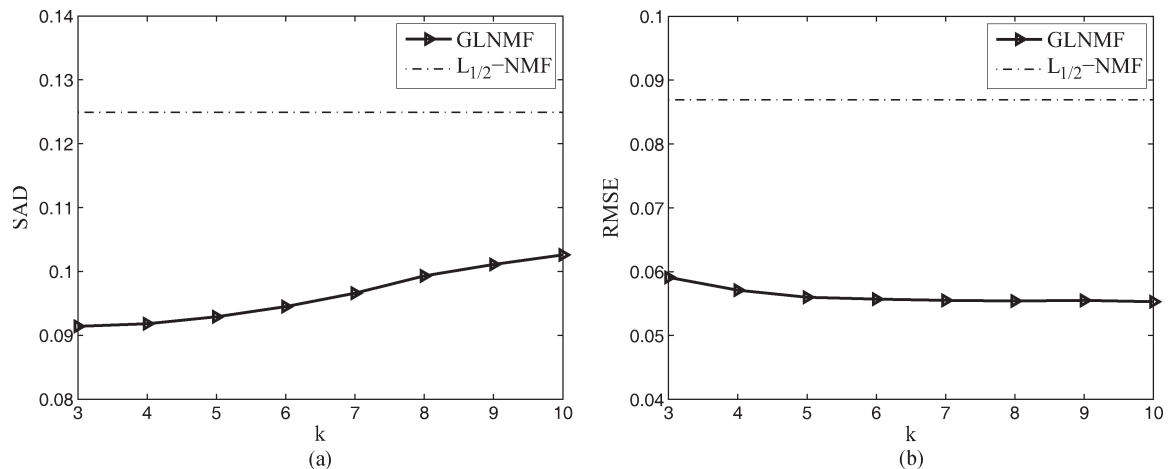
$$\mathbf{A} \leftarrow \mathbf{A} \cdot \mathbf{X} \mathbf{S}^T / \mathbf{A} \mathbf{S} \mathbf{S}^T \quad (14)$$

$$\mathbf{S} \leftarrow \mathbf{S} \cdot (\mathbf{A}^T \mathbf{X} + \mu \mathbf{S} \mathbf{W}) / \left( \mathbf{A}^T \mathbf{A} \mathbf{S} + \frac{1}{2} \lambda \mathbf{S}^{-\frac{1}{2}} + \mu \mathbf{S} \mathbf{D} \right). \quad (15)$$

### C. Implementation Issues

Here, several issues when implementing the algorithm for unmixing will be discussed. In compliance with the process of the proposed algorithm, the first issue is about the initialization methods of the matrices  $\mathbf{A}$  and  $\mathbf{S}$ . Since it is unrealistic to obtain a global minima through optimizing a nonconvex




 Fig. 2. Performance of GLNMF with respect to parameter  $\mu$  in terms of (a) SAD and (b) RMSE.

 Fig. 3. Performance of GLNMF with respect to parameter  $k$  in terms of (a) SAD and (b) RMSE.

function  $f$  in (13), different initial values will bring out different influences. Two initialization methods are studied in this paper: *random initialization* and *VCA-FCLS initialization*. The former randomly chooses values between 0 and 1 as the entries of  $\mathbf{A}$  and  $\mathbf{S}$ . The latter uses VCA to identify endmembers as the input of  $\mathbf{A}$  and then resorts to FCLS to attain the initial  $\mathbf{S}$ . The comparison results of these two initialization methods will be shown in Section IV.

The second important issue is how to ensure ANC and ASC during the iterative process. Considering the updating rules represented by (14) and (15), it is obvious that once the initial values of  $\mathbf{A}$  and  $\mathbf{S}$  are nonnegative, ANC will be strictly guaranteed. In terms of ASC, i.e.,  $\mathbf{1}_P^T \mathbf{S} = \mathbf{1}_N^T$ , a simple but effective technique in [14] is employed here. When updating the matrix  $\mathbf{S}$  by (15), the matrices  $\bar{\mathbf{X}}$  and  $\bar{\mathbf{A}}$  will take the place of  $\mathbf{X}$  and  $\mathbf{A}$  as the inputs, which are defined as

$$\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \delta \mathbf{1}_N^T \end{bmatrix} \quad \bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \delta \mathbf{1}_P^T \end{bmatrix} \quad (16)$$

where  $\delta$  adjusts the effect of ASC. In practice, the larger  $\delta$  can lead to a more accurate result but with a much lower

convergence rate. In order to achieve a desired tradeoff, a relatively small  $\delta = 15$  is selected in the experiments.

Another issue comes from the  $L_{1/2}$  sparsity operator, specifically the term  $\mathbf{S}^{-1/2}$  in (15). Once any zero entry comes into existence in  $\mathbf{S}$ , a very small positive number will be added to make the operation valid. Moreover, as mentioned in [24], those elements in  $\mathbf{S}$  that are less than  $10^{-4}$  will be updated by (15) without their corresponding terms  $(1/2)\lambda \mathbf{S}^{-1/2}$ . By doing this, it can improve the robustness of the algorithm.

Stopping criteria is also a mentionable problem. For the proposed updating rules, a predefined value is set as the error tolerance. If the error does not exceed the tolerance successively for ten times, the iteration ends. In addition, the maximum iteration number is also adopted as a stopping criterion. The number is set to 3000 in the experiments, which is sufficient to guarantee the convergence.

Last, note that how to estimate the number of endmembers has not been mentioned. Although a proper estimation is significant to the result, it is another independent topic, which has not much relation to this paper. One can conduct it by using virtual dimensionality [32] or HySime [33].

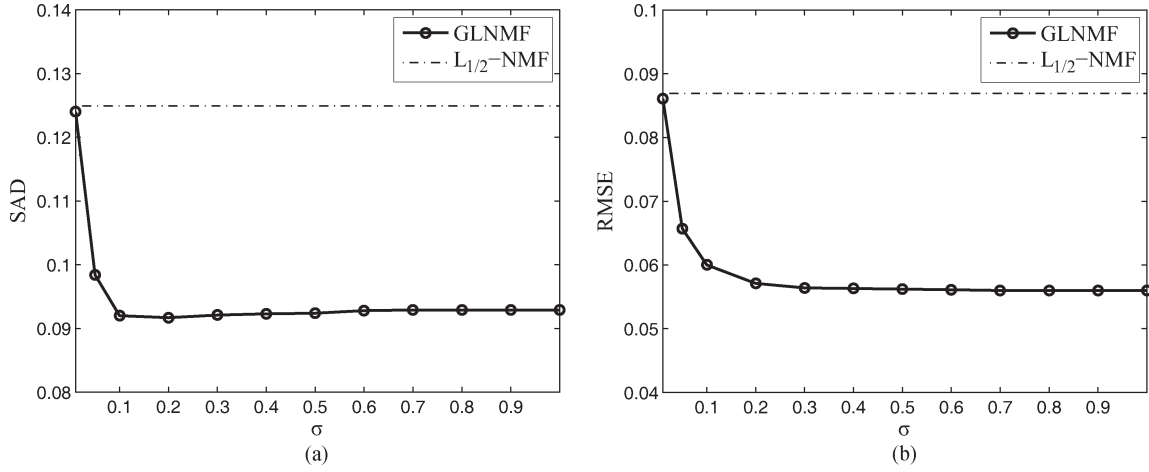


Fig. 4. Performance of GLNMF with respect to parameter  $\sigma$  in terms of (a) SAD and (b) RMSE.

The proposed GLNMF approach (with random initialization) is summarized in Algorithm 1.

---

**Algorithm 1** GLNMF for Hyperspectral Unmixing

**Input:**

The observed mixture data  $\mathbf{X} \in \mathbb{R}^{L \times N}$ ; the number of endmembers  $P$ ; the number of nearest neighbors  $k$ ; and the parameters  $\lambda$ ,  $\mu$ , and  $\sigma$ .

**Output:**

Endmember signature matrix  $\mathbf{A}$  and abundance matrix  $\mathbf{S}$ .

1: **Initialize**  $\mathbf{A}$  and  $\mathbf{S}$  by randomly selecting entries in  $[0, 1]$ . Rescale each column of  $\mathbf{S}$  to unit norm.

2: **Repeat** until convergence:

3: Update  $\mathbf{A}$  by (14);

4: Augment  $\mathbf{X}$  and  $\mathbf{A}$  to obtain  $\bar{\mathbf{X}}$  and  $\bar{\mathbf{A}}$ , respectively;

5: Update  $\mathbf{S}$  by (15).

---

Note that there are some parameters in the algorithm, which may play different roles in the optimization. The parameters  $\lambda$  and  $\mu$  are two essential parameters during the updates.  $\lambda$  has been studied in [24]. Considering that the value of  $\lambda$  relies on the sparseness levels of the material abundances, which are unknown, one possible estimator based on sparseness measure in [34] is given by

$$\lambda = \frac{1}{\sqrt{L}} \sum_l \frac{\sqrt{N} - \|\mathbf{x}_l\|_1 / \|\mathbf{x}_l\|_2}{\sqrt{N} - 1} \quad (17)$$

where  $\mathbf{x}_l$  represents the given hyperspectral data at band  $l$ . For the parameter  $\mu$ , it mainly controls the smoothness level. Thus, one can adjust the ratio  $\lambda/\mu$  to attain a desired result. The other two parameters are the number of nearest neighbors  $k$  and  $\sigma$  in the heat kernel weighting. The selection of these parameters will be concretely shown in the experimental part.

#### D. Computational Complexity Analysis

Here, the computational complexity of the proposed GLNMF approach will be analyzed. It is well known that the computa-

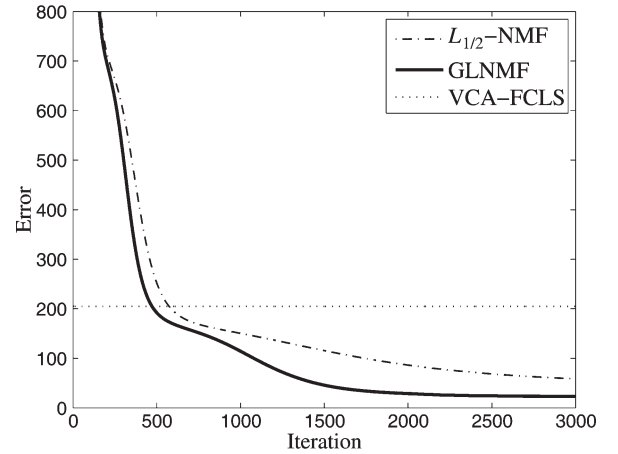


Fig. 5. Comparison of the algorithms in terms of the approximation error.

tional complexity of a standard NMF is  $O(LNP)$ . In order to make the analysis clearer, the time of floating-point calculation has been counted for each iteration.

Concentrating on the rules represented by (14) and (15), it is worth mentioning that the matrix  $\mathbf{W}$  is sparse. There are only  $k$  nonzero elements in each row of  $\mathbf{W}$  on average. Thus, it actually costs  $kPN$  times to do floating-point addition and multiplication in computing  $\mathbf{SW}$ . This is also the case for  $\mathbf{SD}$ . Additionally, the cost of computing  $\mathbf{S}^{-1/2}$  is known as  $(NP)^2$ . Except for this cost, the other three floating-point calculation times for each iteration are listed in Table I.

Note that, at the very beginning, one needs another  $O(N^2L)$  to build the  $k$ -nearest neighbor graph. If the iteration ends after  $m$  steps, the overall cost for a standard NMF is  $O(mLNP)$ . In the GLNMF case, taking all the factors into consideration, the overall cost is  $O(m(L + NP)NP + N^2L)$ .

#### IV. EXPERIMENTS

Here, a series of experiments is designed to evaluate the performance of the proposed GLNMF algorithm. For the synthetic data, GLNMF is mainly compared with the standard NMF,  $L_1$ -NMF [34],  $L_{1/2}$ -NMF, and VCA-FCLS. For the real hyperspectral data, the four counterparts are  $L_{1/2}$ -NMF, VCA,

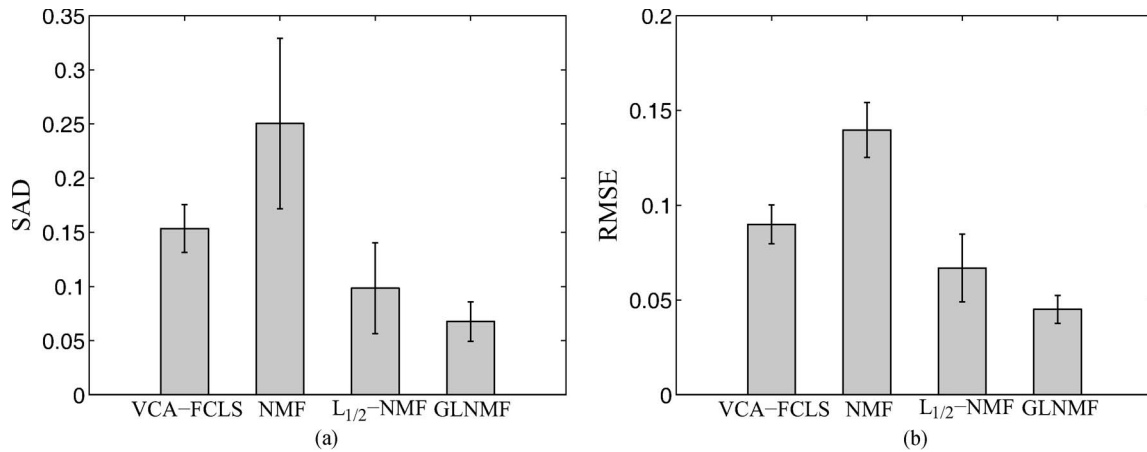


Fig. 6. Comparison of the algorithms in terms of (a) SAD and (b) RMSE.

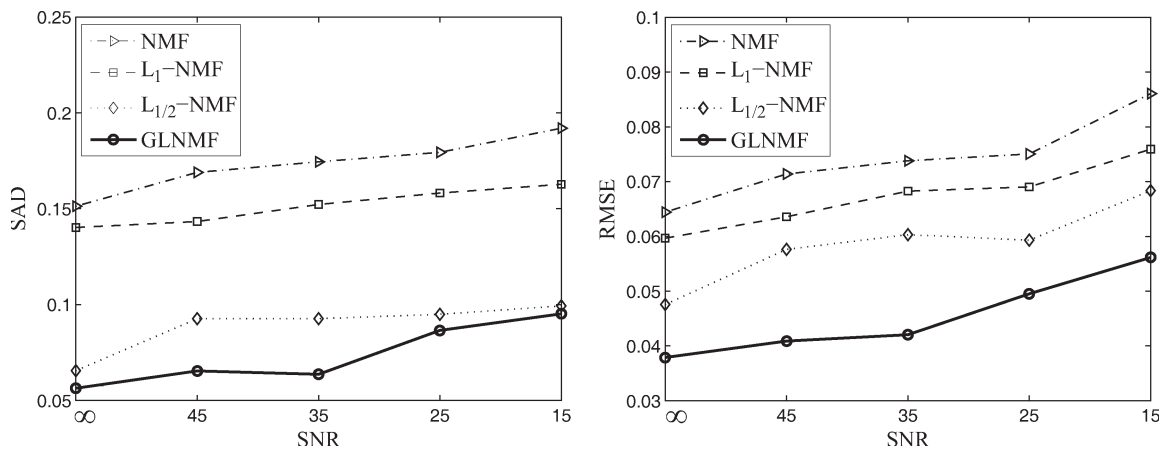


Fig. 7. Comparison of the algorithms at different noise levels in terms of (a) SAD and (b) RMSE.

MVCNMF, and simplex identification via split augmented Lagrangian (SISAL) [35].

Two metrics are used to measure the accuracy of the extracted endmembers and their abundances. They are *spectral angle distance* (SAD) and *root-mean-square error* (RMSE), respectively. For the  $p$ th endmember, suppose the corresponding true signature is  $\hat{\mathbf{A}}_p$ , then the SAD between the estimate  $\hat{\mathbf{A}}_p$  and the true one is defined as

$$\text{SAD}_p = \cos^{-1} \left( \frac{\hat{\mathbf{A}}_p^T \mathbf{A}_p}{\|\hat{\mathbf{A}}_p\| \|\mathbf{A}_p\|} \right). \quad (18)$$

The RMSE is defined as

$$\text{RMSE}_p = \left( \frac{1}{N} |\mathbf{S}_p - \hat{\mathbf{S}}_p|^2 \right)^{\frac{1}{2}} \quad (19)$$

where  $\mathbf{S}_p$  and  $\hat{\mathbf{S}}_p$  denote the estimated abundance matrix and the ground-truth abundance matrix, respectively.

#### A. Experiments on Synthetic Data

To generate the synthetic data, ten spectra are randomly chosen from the USGS digital spectral library [36]. Among them, the first six true signatures are shown in Fig. 1. Then, the generation process similar to that in [22] can be described

as follows: 1) An image of size  $64 \times 64$  is divided into  $8 \times 8$  patches; 2) each patch is filled up by only one type of signature, which is randomly selected from the ten; 3) a  $7 \times 7$  low-pass filter is utilized to generate mixed pixels; and 4) for the pixel whose abundance is larger than 0.8, a mixture composed of all endmembers with the abundances of  $1/P$  will take the place of it.

By following these procedures, a synthetic image is produced without the existence of pure pixels. Note that the low-pass filter not only makes the pixels highly mixed but also makes the abundance vary smoothly, which conforms to the real-world case. Last, to evaluate the performance of resistance to noise, zero-mean white Gaussian noise is added into the synthetic data. Here, the *signal-to-noise ratio* (SNR) is defined as

$$\text{SNR} = 10 \log_{10} \frac{E[\mathbf{x}^T \mathbf{x}]}{E[\mathbf{e}^T \mathbf{e}]} \quad (20)$$

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

1) *Experiment 1 (Parameters Selection)*: In this experiment, the selections of three parameters  $\mu$ ,  $k$ , and  $\sigma$  are considered when  $\text{SNR} = 15$  dB,  $P = 4$ , and  $\lambda = 0.1$ . To eliminate the effect of random initialization, the same initial  $\mathbf{A}$  and  $\mathbf{S}$  are adopted in both  $L_{1/2}$ -NMF and GLNMF. Here, the result of  $L_{1/2}$ -NMF is only taken as a reference. Fig. 2 shows how

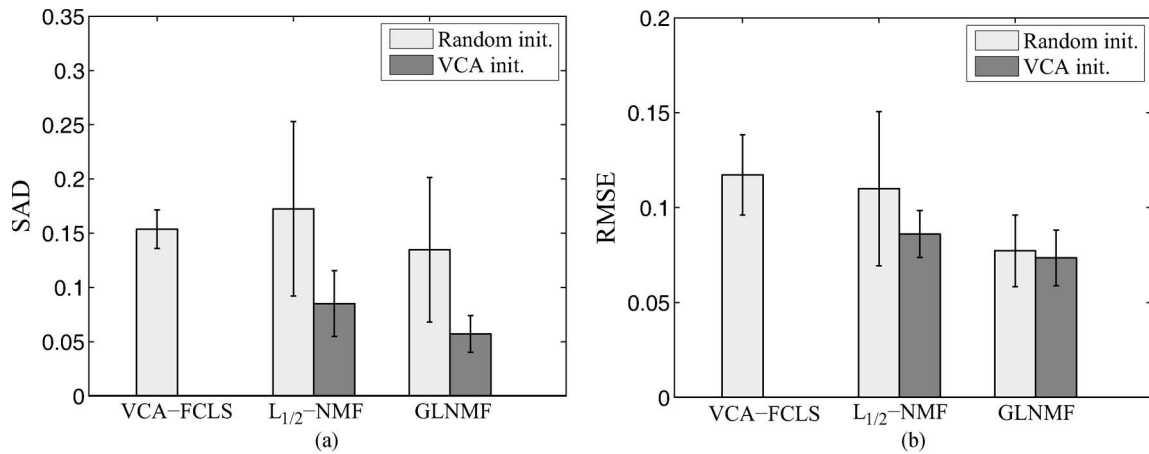


Fig. 8. Comparison of the algorithms with different initialization methods in terms of (a) SAD and (b) RMSE.

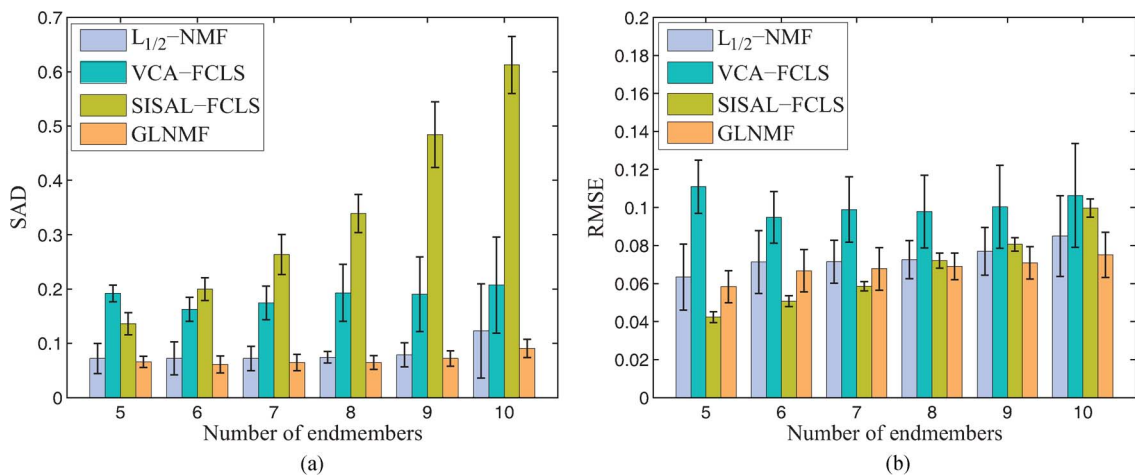


Fig. 9. Comparison of the algorithms with different numbers of endmembers in terms of (a) SAD and (b) RMSE.

the performance varies with the parameter  $\mu$  when the parameters are set as  $k = 5$  and  $\sigma = 1$ . As shown, GLNMF attains constantly good performance in terms of RMSE when  $\mu$  varies from 0.1 to 0.4. It further verifies the effectiveness of incorporating the graph regularization term, which can provide a better abundance estimation. On the other hand, there is an increase in SAD during this interval. To achieve a desired result, the parameter  $\mu$  is set to 0.1 in the experiments. When fixing  $\mu = 0.1$  and  $\sigma = 1$ , Fig. 3 gives the performance variation with respect to the number of nearest neighbors  $k$ . Likewise, the RMSE of GLNMF is very stable when  $k$  changes from 3 to 10, whereas the SAD slowly increases. Thus,  $k = 5$  can be a relatively good choice. For the parameter  $\sigma$ , Fig. 4 demonstrates the experimental results. It is worth noting that when  $\sigma$  goes to 0.3, the RMSE and SAD of GLNMF seem not to change anymore. Recall (12), the heat kernel weighting will surely approximate to 0–1 weighting when  $\sigma$  is large enough. Therefore, both of the two curves arrive at a stationary state. In the experiments, the parameter  $\sigma$  is selected as  $\sigma = 1$ .

2) *Experiment 2 (Convergence Analysis)*: Under the condition of the former experiment, the approximation errors of VCA-FCLS,  $L_{1/2}$ -NMF, and GLNMF are plotted in Fig. 5. Since VCA-FCLS is deterministic, the approximation error of VCA-FCLS keeps the same, whereas the convergence

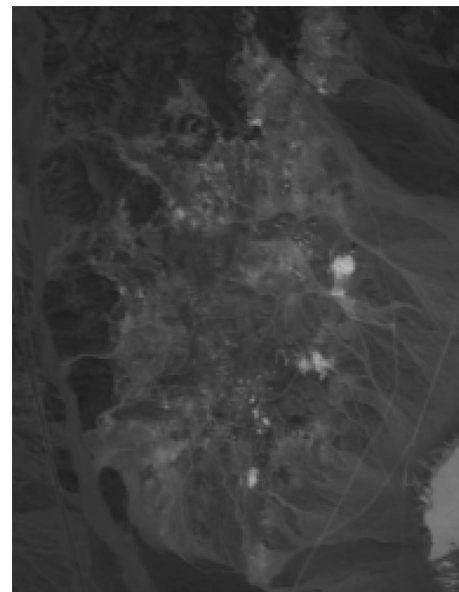


Fig. 10. Band 30 of the Cuprite image.

curve of GLNMF declines faster than that of  $L_{1/2}$ -NMF, and GLNMF also yields the smallest error among the three methods.



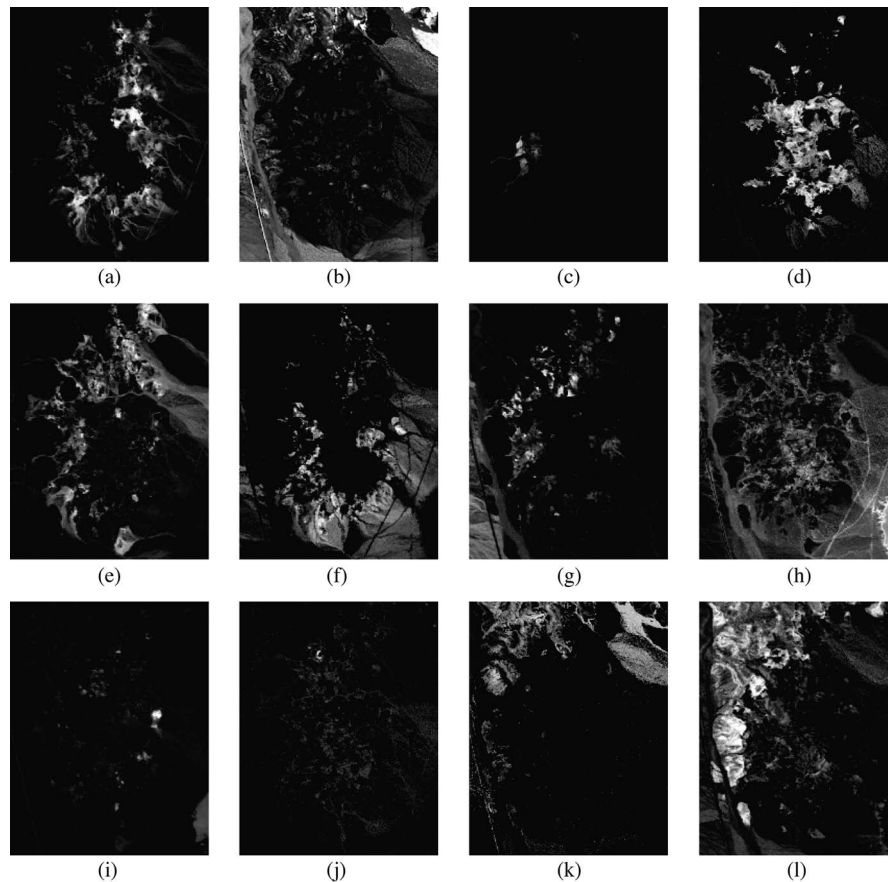


Fig. 11. Abundance maps of different endmembers using GLNMF on the AVIRIS Cuprite data. (a) Alunite. (b) Andradite. (c) Buddingtonite. (d) Chalcedony. (e) Dumortierite. (f) Kaolinite #1. (g) Kaolinite #2. (h) Montmorillonite. (i) Muscovite. (j) Nontronite. (k) Pyrope. (l) Sphene.

3) *Experiment 3 (Comparison of Different Algorithms)*: This experiment aims at evaluating the unmixing performances of VCA-FCLS, NMF,  $L_{1/2}$ -NMF, and GLNMF. For a fair comparison, the standard NMF also takes the technique in (16) when updating  $\mathbf{S}$ , and all the NMF-based algorithms adopt random initialization. Fig. 6 shows the results when  $\text{SNR} = 20$  dB and  $P = 4$ , where the bar and the error bar stand for the mean and the standard deviation, respectively. It is clear that GLNMF leads to the best estimation in both endmembers and abundances. Moreover, NMF and  $L_{1/2}$ -NMF are more easily affected by the random initialization because they show larger deviations compared with the other two.

4) *Experiment 4 (Robustness Analysis to Noise)*: In this experiment, four NMF-based algorithms are performed with different noise levels when  $P$  is set to 4. The SNR is assigned to 15 dB, 25 dB, 35 dB, 45 dB, and infinity (noise-free). Since NMF and  $L_1$ -NMF are not specialized to unmixing, the technique in (16) is also employed here, and all the algorithms take random initialization. Fig. 7(a) demonstrates that  $L_1$ -NMF slightly outperforms NMF when the SNR varies, and GLNMF still results in the smallest SAD. For the RMSE, GLNMF also turns out to be the best, as shown in Fig. 7(b). Particularly when  $\text{SNR} = 15$  dB, the SAD of GLNMF is similar to that of  $L_{1/2}$ -NMF, but the RMSE is obviously superior to that of  $L_{1/2}$ -NMF. It reveals that GLNMF can contribute to a better estimate of abundances.

5) *Experiment 5 (Impact of Different Initializations)*: This experiment shows what influences would be produced by different initialization methods under the condition of  $\text{SNR} = 10$  dB and  $P = 4$ . As shown in Fig. 8, when adopting random initialization, both  $L_{1/2}$ -NMF and GLNMF give relatively large standard deviation compared with VCA. However, when adopting VCA-FCLS initialization, the standard deviation apparently decreases, and both the SAD and the RMSE improve. It is mainly because VCA-FCLS initialization provides a more accurate initial point, which is closer to the global optimum. This reduces the potential of being trapped in many local minima during the iterative procedure.

6) *Experiment 6 (Generalization to Different Numbers of Endmembers)*: This experiment evaluates the performances of VCA-FCLS,  $L_{1/2}$ -NMF, SISAL-FCLS, and GLNMF when the data are constituted of different numbers of endmembers. When  $\text{SNR} = 20$  dB and  $P$  varies from 5 to 10, the results are shown in Fig. 9. Overall, the performances decay when the number increases. VCA fails to obtain the most accurate results due to the lack of pure pixels. Meanwhile, the performance of SISAL decreases when the number of endmembers increases. It is mainly due to the noise, which strongly affects the volume of the minimum enclosing simplex. As the number of endmembers increases, the influence becomes more significant. GLNMF outperforms  $L_{1/2}$ -NMF a little when  $P = 5$ , but the superiority becomes larger as  $P$  increases.

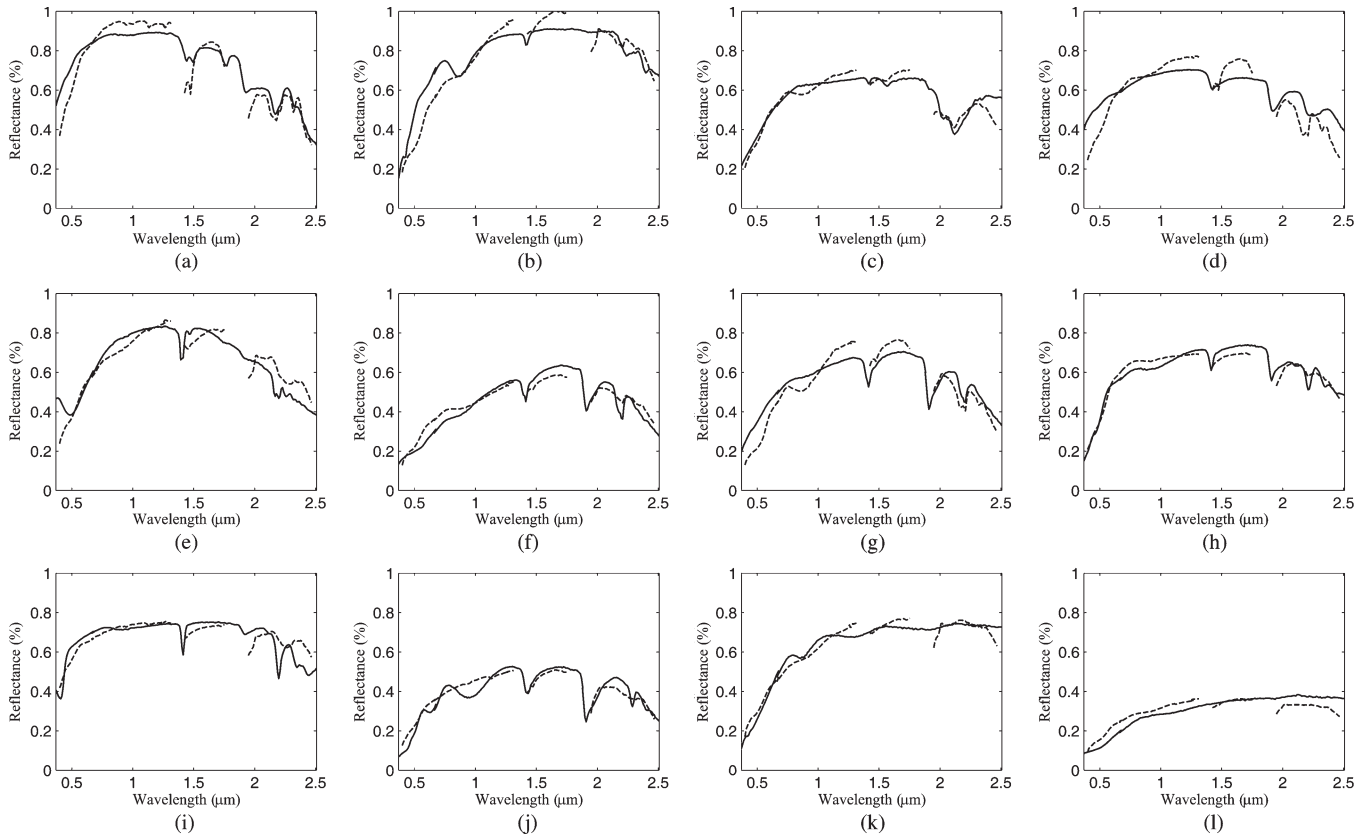


Fig. 12. Comparison of the USGS library spectra (solid line) with the endmember signatures extracted by GLNMF (dotted line) on the AVIRIS Cuprite data. (a) Alunite. (b) Andradite. (c) Buddingtonite. (d) Chalcedony. (e) Dumortierite. (f) Kaolinite #1. (g) Kaolinite #2. (h) Montmorillonite. (i) Muscovite. (j) Nontronite. (k) Pyrope. (l) Sphene.

TABLE II  
SAD AND THE STANDARD VARIANCE (%) RESULTS ON THE AVIRIS CUPRITE DATA

	GLNMF	$L_{1/2}$ -NMF	VCA	MVCNMF	SISAL
Alunite	$0.1170 \pm 5.02$	<b><math>0.1039 \pm 3.67</math></b>	$0.1129 \pm 5.49$	$0.1453 \pm 4.24$	$0.2054 \pm 2.74$
Andradite	<b><math>0.0731 \pm 1.68</math></b>	$0.0822 \pm 2.86$	$0.0886 \pm 3.75$	$0.0971 \pm 4.52$	$0.0992 \pm 2.91$
Buddingtonite	<b><math>0.0759 \pm 0.62</math></b>	$0.0998 \pm 1.52$	$0.0900 \pm 1.94$	$0.1143 \pm 2.82$	$0.1217 \pm 1.65$
Chalcedony	$0.1438 \pm 2.08$	$0.1307 \pm 2.10$	$0.1296 \pm 3.71$	<b><math>0.1162 \pm 5.12</math></b>	$0.1316 \pm 2.33$
Dumortierite	<b><math>0.0778 \pm 1.12</math></b>	$0.0932 \pm 3.00$	$0.0861 \pm 1.46$	$0.1311 \pm 4.67$	$0.1247 \pm 1.89$
Kaolinite #1	$0.0851 \pm 1.46$	<b><math>0.0778 \pm 0.77</math></b>	$0.0790 \pm 1.30$	$0.0871 \pm 2.96$	$0.1441 \pm 3.41$
Kaolinite #2	$0.0853 \pm 2.95$	$0.0692 \pm 2.30$	$0.0704 \pm 1.77$	$0.1044 \pm 5.25$	<b><math>0.0609 \pm 1.46</math></b>
Montmorillonite	$0.0848 \pm 3.88$	$0.0695 \pm 2.61$	$0.0637 \pm 3.12$	$0.0922 \pm 5.58$	<b><math>0.0636 \pm 2.39</math></b>
Muscovite	<b><math>0.1025 \pm 3.59</math></b>	$0.1644 \pm 2.89$	$0.1443 \pm 4.31$	$0.1568 \pm 5.18$	$0.1336 \pm 2.27$
Nontronite	$0.0902 \pm 1.14$	$0.0894 \pm 1.59$	<b><math>0.0861 \pm 1.73</math></b>	$0.0947 \pm 3.25$	$0.1023 \pm 2.38$
Pyrope	<b><math>0.0596 \pm 0.56</math></b>	$0.0616 \pm 0.83$	$0.0861 \pm 5.61$	$0.0701 \pm 1.82$	$0.1580 \pm 3.26$
Sphene	$0.1467 \pm 4.50$	$0.1628 \pm 6.29$	$0.1368 \pm 8.18$	<b><math>0.0654 \pm 1.85</math></b>	$0.2128 \pm 3.38$
Mean	<b><math>0.0951 \pm 2.38</math></b>	$0.1004 \pm 2.54$	$0.0978 \pm 3.53$	$0.1062 \pm 3.94$	$0.1298 \pm 2.51$

### B. Experiments on Real Hyperspectral Data

Here, GLNMF is applied into real hyperspectral image collected by the *Airborne Visible/Infrared Imaging Spectrometer* (AVIRIS) over Cuprite, Nevada. These data have been widely used for assessing unmixing algorithms [10], [37], [38]. Fig. 10 shows the 30th band of the original image with size  $250 \times 191$ . In the Cuprite data, there are 224 bands that cover the wavelength range of  $0.4\text{--}2.5\text{ }\mu\text{m}$ . Before unmixing, low SNR and water vapor absorption bands (including bands 1–2, 104–113, 148–167, and 221–224) are removed, leaving 188 bands in the experiment.

In the light of the existing analysis in [10], 14 types of minerals are estimated in the image. Since there is a subtle difference among some spectra of the same mineral with different chemical compositions, the estimated number of endmembers is reduced to 12 for unmixing. Fig. 11 illustrates the resulting abundances by GLNMF. It is a grayscale abundance map, where a dark pixel denotes low abundance of the relative endmember. Meanwhile, Fig. 12 shows the comparison between the extracted endmember signatures by GLNMF and the corresponding USGS library spectra, and Table II gives the quantitative result in terms of the SAD. It is shown that GLNMF

obtains the best performance for most of the minerals, yielding the lowest mean and variance.

## V. CONCLUSION

In this paper, a new hyperspectral unmixing algorithm, which utilizes the intrinsic structural information, has been proposed. By virtue of introducing the manifold regularization into the sparsity-constrained NMF, the proposed GLNMF can lead to a more desirable result. Compared with other state-of-the-art algorithms, the superiority of GLNMF mainly depends on preserving the close link between the observation and the endmember abundances by incorporating the manifold regularization term. In addition, GLNMF owns a better antinoise capability. The experimental analyses further indicate the validity of GLNMF, which can provide more accurate estimates.

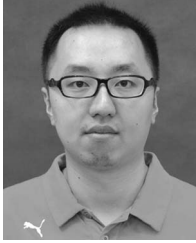
## REFERENCES

- [1] A. Goetz, G. Vane, J. Solomon, and B. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, Jun. 1985.
- [2] G. Foody, *Remote Sensing Image Analysis: Including the Spatial Domain*. Norwell, MA: Kluwer, 2004, ch. 3, pp. 37–49.
- [3] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [4] A. Plaza, Q. Du, J. Bioucas-Dias, X. Jia, and A. Kruse, "Foreword to the special issue on spectral unmixing of remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 1–8, Nov. 2011.
- [5] T. Lillesand and R. Kiefer, *Remote Sensing and Image Interpretation*, 4th ed. New York: Wiley, 2000.
- [6] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [7] X. Ceamanos, S. Douté, B. Luo, F. Schmidt, G. Jouannic, and J. Chanussot, "Intercomparison and validation of techniques for spectral unmixing of hyperspectral images: A planetary case study," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4341–4358, Nov. 2011.
- [8] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summ. 4th Annu. JPL Airborne Geosci. Workshop*, R. O. Green, Ed., 1994, pp. 11–14.
- [9] M. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Conf. Imag. Spectrom.*, V, 1999, pp. 266–275.
- [10] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [11] J. Bayliss, J. Gualtieri, and R. Cromp, "Analyzing hyperspectral data with independent component analysis," in *Proc. SPIE*, 1997, vol. 3240, pp. 133–143.
- [12] J. Nascimento and J. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005.
- [13] W. Xia, X. Liu, B. Wang, and L. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2165–2179, Jun. 2011.
- [14] D. Heinz and C. 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.
- [15] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [16] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.
- [17] D. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng, "Integration of spatial-spectral information for the improved extraction of endmembers," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 287–303, 2007.
- [18] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2679–2693, Aug. 2009.
- [19] C. Liou and K. Yang, "Unsupervised classification of remote sensing imagery with non-negative matrix factorization," in *Proc. ICONIP*, 2005, pp. 280–285.
- [20] P. Hoyer, "Non-negative sparse coding," in *Proc. 12th IEEE Workshop NNISP*, 2002, pp. 557–565.
- [21] V. Pauca, J. Piper, and R. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.
- [22] 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.
- [23] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.
- [24] 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.
- [25] Z. Xu, H. Zhang, Y. Wang, and Y. Chang, " $L_{1/2}$  regularizer," *Sci. China. Ser. F. Inf. Sci.*, vol. 53, no. 6, pp. 1159–1169, Jun. 2010.
- [26] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 585–591, 2001.
- [27] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2001, pp. 556–562.
- [28] L. Ma, M. Crawford, and J. Tian, "Local manifold learning-based  $k$ -nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2009.
- [29] J. He, L. Zhang, Q. Wang, and Z. Li, "Using diffusion geometric coordinates for hyperspectral imagery representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 4, pp. 767–771, Oct. 2009.
- [30] M. Hein and M. Maier, "Manifold denoising," *Adv. Neural Inf. Process. Syst.*, vol. 20, pp. 1–8, 2006.
- [31] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [32] C. 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.
- [33] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [34] P. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.
- [35] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE Workshop Hyperspectral Image Signal Process.—Evol. Remote Sens.*, 2009, pp. 1–4.
- [36] R. Clark, G. Swayze, A. Gallagher, T. King, and W. Calvin, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," U.S. Geol. Surv., Denver, CO, 1993, Open File Rep..
- [37] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 757–772, Feb. 2011.
- [38] M. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.



**Xiaoqiang Lu** received the Ph.D. degree from Dalian University of Technology, Dalian, China, in 2010.

He is a Postdoctoral Researcher with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. His research interests include pattern recognition, machine learning, hyperspectral image analysis, cellular automata, and medical imaging.



**Hao Wu** received the B.S. degree from Zhejiang University, Hangzhou, China, in 2009. He is currently working toward the M.S. degree in the Faculty of Mathematics and Computer Science, Hubei University, Wuhan, China.

His main research interests include hyperspectral image analysis, pattern recognition, and machine learning.



**Pingkun Yan** (S'04–M'06–SM'10) received the B.Eng. degree in electronics engineering and information science from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore.

He is currently a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. His research

interests include computer vision, pattern recognition, machine learning, and their applications in medical imaging.

**Yuan Yuan** (M'05–SM'09) is a Researcher (Full Professor) with the Chinese Academy of Sciences. Her main research interests include visual information processing and image/video content analysis.

**Xuelong Li** (M'02–SM'07–F'12) is currently a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.