

# Blind Spectral Unmixing Based on Sparse Nonnegative Matrix Factorization

Zuyuan Yang, Guoxu Zhou, Shengli Xie, *Senior Member, IEEE*, Shuxue Ding, *Member, IEEE*, Jun-Mei Yang, and Jun Zhang

**Abstract**—Nonnegative matrix factorization (NMF) is a widely used method for blind spectral unmixing (SU), which aims at obtaining the endmembers and corresponding fractional abundances, knowing only the collected mixing spectral data. It is noted that the abundance may be sparse (i.e., the endmembers may be with sparse distributions) and sparse NMF tends to lead to a unique result, so it is intuitive and meaningful to constrain NMF with sparseness for solving SU. However, due to the abundance sum-to-one constraint in SU, the traditional sparseness measured by L0/L1-norm is not an effective constraint any more. A novel measure (termed as S-measure) of sparseness using higher order norms of the signal vector is proposed in this paper. It features the physical significance. By using the S-measure constraint (SMC), a gradient-based sparse NMF algorithm (termed as NMF-SMC) is proposed for solving the SU problem, where the learning rate is adaptively selected, and the endmembers and abundances are simultaneously estimated. In the proposed NMF-SMC, there is no pure index assumption and no need to know the exact sparseness degree of the abundance in prior. Yet, it does not require the preprocessing of dimension reduction in which some useful information may be lost. Experiments based on synthetic mixtures and real-world images collected by AVIRIS and HYDICE sensors are performed to evaluate the validity of the proposed method.

**Index Terms**—Blind spectral unmixing, nonnegative matrix factorization (NMF), sparseness measure.

## I. INTRODUCTION

WITH the rapid development of the space technology, remote sensing image attracts more and more attention for its great potential civilian and military applications, such as mineral exploration, environmental monitoring, military surveillance, etc. [1], [2]. When the spatial resolution of the imaging sensor is low or the distinct materials are combined into a homogeneous mixture (e.g., sand grains on a beach), there are

Manuscript received February 16, 2009; revised October 20, 2009, March 15, 2010, June 23, 2010; accepted September 18, 2010. Date of publication September 30, 2010; date of current version March 18, 2011. This work is supported by National Basic Research Program of China (973 Program, No. 2010CB731800), National Natural Science Foundation of China under Grants U0635001, U0835003, 60874061, 60974072, and in part by the 2008–2010 Grants-In-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science and in part by Technology, Japan, Project No. 20500209. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Birsen Yazici.

Z. Y. Yang, G. X. Zhou, S. L. Xie, and J. M. Yang are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510641, China (e-mail: eoshlxie@scut.edu.cn).

S. Ding is with the School of Computer Science and Engineering, the University of Aizu, Tsuruga, Fukushima 965-8580, Japan, and also with the Brain Science Institute, RIKEN, Saitama 351-0198, Japan (e-mail: sdng@u-aizu.ac.jp).

J. Zhang is with Beihang University, Beijing 100191, China (e-mail: buazhangjun@vip.sina.com).

Digital Object Identifier 10.1109/TIP.2010.2081678

many mixed pixels in the remote sensing image [3]. In this case, the measured spectrum of a single pixel is a mixture of several endmember spectra, weighted by their fractional abundances. How to decompose the mixed pixels is a common problem, that is so-called spectral unmixing (SU) [4], [5]. The purpose of SU is to recover the endmembers and corresponding abundances, knowing only the collected mixed spectral data. Mathematically, this problem corresponds to the typical blind source separation (BSS), where the collected mixtures, endmember signatures and abundances in SU can be considered as the observations, mixing matrix and sources in BSS, respectively [6]–[8].

For solving the SU/BSS problem mentioned above, there are mainly two classes of methods: 1) the two-stage methods (TSM), which extract or separate endmembers firstly, then, estimate the abundance based on the extracted endmembers. There are several ways for endmember extraction, such as the local maximum-likelihood algorithm based on a subspace model [9], the spatial preprocessing method [10], the maximum-likelihood algorithm based on Gauss-Markov random field modeling [11], the support vector machine (SVM) based method [12], [13], the convex geometry optimization methods based on the pure pixel assumption, e.g., N-FINDR [14], the pixel purity index (PPI) [15], [16], and the vertex component analysis (VCA) [17], etc. The abundance is mainly estimated by the fully constrained least square (FCLS), subject to full additivity and positivity constraints [18], the artificial neural networks [19], the fuzzy c-means clustering [20], etc. Moreover, a standardized processing method for hyperspectral data analysis is developed by Kruse *et al.* [21], [22]. 2) the single stage methods (SSM), which obtain the endmembers and their abundances simultaneously, such as the unsupervised FCLS (UFCLS) [23], the gradient descent maximum entropy (GDME) approach [24], independent component analysis (ICA) [7], the maximum non-Gaussianity (MaxNG) [8], spatial complexity BSS (SCBSS) method [25], etc. Reference [3] shows many techniques for hyperspectral image unmixing, including the hierarchical taxonomies that reveal the commonalities and differences between unmixing algorithms.

In TSM, the endmembers and abundances are estimated subsequently. If the estimation precision of the endmembers is low, it may lead to a large cumulative error in the estimation of the abundances. Yet, each of existing SSM has its specific limitation in applications. ICA is widely used for BSS, under the assumption that sources are independent. However, due to the abundance sum-to-one constraint (ASC) in SU, the application of ICA is restricted [7], [8]. MaxNG gets out of the independence of abundances, but it is computationally expensive [8].

UFCLS is a computationally efficient method, but it focuses on least-square optimization, which could suffer from the overfitting problem [23], [24]. Recently, based on the classic maximum entropy principle, GDME has been shown working effectively for solving SU, but the learning rate is not adaptively selected in this gradient based method [24]. Furthermore, by using the spatial complexity, SCBSS estimates endmembers and abundances with a low cost. However, the unmixing results may contain some negative values which are meaningless obviously [25].

On the other hand, nonnegative matrix factorization (NMF) can obtain nonnegative results with physical significance [26]. Therefore, it has great potential for solving SU [27]–[31]. However, it shows that NMF itself does not necessarily provide immediate interpretation and therefore some additional constraints on the factors are needed for NMF applications [32]. In fact, many methods based on constrained NMF have been developed for solving SU. For instance, NMF based on minimum volume constraint (NMF-MVC) can solve mixed spectral data using convexity of mixtures [1]. However, it requires the pre-processing of dimensionality reduction in which some useful information may be lost. In [33], NMF based on minimum distance constraint is utilized to avoid numerical instabilities in NMF-MVC, where the volume-based constraint is replaced by distance-based constraint. Notably it is an iterated algorithm and the abundances are estimated using least-squares based method in each step. Therefore, it suffers from the overfitting problem. Furthermore, NMF based on smoothness constraint (NMF-SC) is utilized to analyze spectral data [27]. As the collected dataset is normalized in a preprocessing step and the endmember signatures are scaled to sum-to-one at each iteration, the abundance is hard to be estimated only using NMF-SC [27]. In fact, NMF-SC is mainly used to estimate the endmembers (the abundances are generally estimated using another scheme, such as FCLS [18], artificial neural networks [19], etc). Recently, the piecewise smooth NMF with sparseness constraint (PSNMFSC) algorithm is proposed for spectral unmixing [34]. Due to the united constraints of the piecewise smoothness and the sparseness, PSNMFSC performs well. However, it often needs some prior knowledge about the exact sparseness degree of the abundances.

This paper will focus on the sparseness constraint to the abundances for solving SU by NMF. Actually, due to the optimization is simple and the results tend to be unique, sparse NMF itself is always a hot topic [35]–[37]. In order to solve SU by using sparse NMF efficiently, a new sparseness measure (S-measure) is proposed (see Section III.A). It takes advantage of higher-order norms of signal vectors and features physical significance. Furthermore, by using the S-measure constraint (SMC), a gradient-based sparse NMF (termed as NMF-SMC) is proposed to estimate the endmembers and abundances simultaneously, where the learning rate is adaptively selected. There is no pure pixel assumption in NMF-SMC yet it needs not a preprocessing of dimension reduction which may make some useful information be lost.

The rest of this paper is organized as follows. In Section II, typical SU and NMF models are presented. Also, the physical features of endmembers and abundances are analyzed. Section III describes the sparse NMF, in addition with the new sparseness

measure and the gradient based optimization algorithm NMF-SMC. The performance evaluations using synthetic mixtures and real hyperspectral images are presented in Section IV. Finally, conclusions are summarized briefly in Section V.

## II. SPECTRAL UNMIXING AND NMF

Generally speaking, both of linear models and nonlinear models are proper for SU, depending on the distribution of endmembers [38], [39]. When endmembers are distributed as discrete patches and different endmembers do not interfere with each other, the linear model is effective [1], [24]. For simplicity, only linear model is considered in this paper.

### A. Linear SU Model

Typical linear SU model is as follows:

$$\mathbf{V} = \mathbf{WH} + \boldsymbol{\varepsilon} \quad (1)$$

where  $\mathbf{V} = [\mathbf{v}_1^T, \dots, \mathbf{v}_m^T]^T \in \mathbb{R}^{m \times n}$  denotes the collected mixtures,  $\mathbf{W} = [\mathbf{w}_1^T, \dots, \mathbf{w}_r^T]^T \in \mathbb{R}^{m \times r}$  denotes the endmember signatures,  $\mathbf{H} = [\mathbf{h}_1^T, \dots, \mathbf{h}_n^T]^T \in \mathbb{R}^{r \times n}$  denotes the corresponding abundances, with  $m, r, n$  denoting the number of bands in endmember signatures, the number of endmembers and the number of pixels in remote sensing image, respectively. Both  $\mathbf{W}$  and  $\mathbf{H}$  are unknown. And  $\boldsymbol{\varepsilon}$ , the possible errors or noises, is also unknown.

Based on the real background, it is found that  $\mathbf{W}, \mathbf{H}$  in the model (1) have the following features:  $\mathbf{W}, \mathbf{H}$  are nonnegative, and each column of  $\mathbf{H}$  is sum to one [1], [23], [24]. Furthermore,  $\mathbf{H}$  may have sparseness feature [34], [40]. Just like what Jia and Qian said, in most cases, any endmember does not always contribute to all pixels in the scene, i.e., the abundance of each endmember is located [34]. Actually, many existing methods for solving SU implicate the sparseness of the abundances. For instance, clustering algorithm reflects the sufficient sparseness of  $\mathbf{H}$  [12], [20]; In convex hull geometry algorithm, such as N-FINDR, PPI, and VCA, the abundance corresponding to pure pixel is sufficiently sparse [14], [15], [17]; NMF-MVC relies on the fact that the simplex volume determined by the endmembers is the minimum among all possible simplexes that circumscribe the data scatter space [1]. To determine the end-members, it needs the assumption that there exist many points distributed on the bounds, which fully reflects the sparseness of the abundances [41], [42], etc.

As only  $\mathbf{V}$  is known in the model (1), this linear SU model is equivalent to the instantaneously linear BSS model in [6], where  $\mathbf{V}, \mathbf{W}, \mathbf{H}$  are used to denote the mixed signals, mixing matrix and sources, respectively [6]–[8]. Since  $\mathbf{W}, \mathbf{H}$  are both nonnegative, it is natural to use NMF based BSS for solving the model (1) [27]–[29].

### B. NMF

Considering only the nonnegativity constraint, we can convert the problem of solving the model (1) into optimizing the following model [26], [43]:

$$\begin{aligned} \text{Min : } D(\mathbf{W}, \mathbf{H}) &= \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_2^2 \\ \text{s.t. } w_{ij} &\geq 0, h_{jt} \geq 0, \forall i \in \mathcal{M}, j \in \mathcal{R}, t \in \mathcal{N} \end{aligned} \quad (2)$$

where  $\mathcal{M} = \{1, \dots, m\}$ ,  $\mathcal{R} = \{1, \dots, r\}$ , and  $\mathcal{N} = \{1, \dots, n\}$  (throughout this paper).

There are several methods for solving the above problem [43], [44]. The most popular one is the alternate-least-squares (ALS) gradient-based updating algorithm. ALS begins with two initial positive matrices  $\mathbf{W}, \mathbf{H}$ , and then updates them at each iteration using the following scheme: fixing one matrix while optimizing the other. Under the ALS gradient-based updating rule, the cost function in the model (2) is monotonically non-increasing [43].

Due to the permutation and scale indeterminancies which widely exist in BSS (i.e., for permutation matrix  $\mathbf{P}$  and diagonal matrix  $\mathbf{D}$ , it holds  $\mathbf{V} = \mathbf{WH} = (\mathbf{WP}^{-1}\mathbf{D}^{-1})(\mathbf{DPH})$ ), NMF itself does not necessarily provide immediate meaningful interpretation, therefore, some additional constraints on the factors are needed [32]. Constrained NMF model is generally given as follows [45]:

$$\begin{aligned} \text{Min : } D_\alpha &= \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_2^2 + \alpha_{\mathbf{W}} J_{\mathbf{W}}(\mathbf{W}) + \alpha_{\mathbf{H}} J_{\mathbf{H}}(\mathbf{H}) \\ \text{s.t. } w_{ij} &\geq 0, h_{jt} \geq 0, \forall i \in \mathcal{M}, j \in \mathcal{R}, t \in \mathcal{N} \end{aligned} \quad (3)$$

where  $\alpha_{\mathbf{W}}, \alpha_{\mathbf{H}} \geq 0$  are regularization parameters,  $J_{\mathbf{W}}, J_{\mathbf{H}}$  are measure functions (rely on the expected constraints) for  $\mathbf{W}$  and  $\mathbf{H}$ , respectively. As has been shown that the matrix  $\mathbf{H}$  may be sparse besides the normal nonnegativity and sum-to-one (see Section II.A), the sparseness constraint to  $\mathbf{H}$  will be focused and discussed in detail in this paper (i.e.,  $\alpha_{\mathbf{W}} = 0$ ).

### III. SPARSENESS CONSTRAINT BASED NMF

NMF based on sparseness constraint will be discussed and used to solve SU in this section. The sparseness of the abundance matrix  $\mathbf{H}$  is a widely referred feature, but few algorithms utilize it to solve SU directly. One reason is that it is hard for how to measure sparseness properly (i.e., it is hard to select the function  $J_{\mathbf{H}}$  in the model (3)). In fact, as argued in [46], the selection of the sparseness measure is not a minor technical detail but may have far-reaching implications on the structure of a solution. A novel sparseness measure will be developed in this section.

#### A. Sparseness Measure

Traditionally, L0-norm has been used as the sparseness measure [42], [47]. Since the corresponding measure function  $J_{\mathbf{H}}$  is not differentiable, L0-norm is often replaced by L1-norm in real applications (because the optimized solutions are similar by using these two norms) [48], [49]. However, due to ASC in SU, it is not proper to use L1-norm to construct the measure function  $J_{\mathbf{H}}$  directly. Some nonlinear functions have also been used in  $J_{\mathbf{H}}$ , such as  $\log(1+x^2)$  and  $-\exp(-x^2)$  in [50],  $\tanh(|ax|^b)$  in [51], and generalized Gaussian probability density function (PDF)  $(\alpha)/(2\beta\Gamma(1/\alpha))\exp(-(|x|/\beta)^\alpha)$  in [52]. The generalized Gaussian PDF can feature physical significance because it reflects the distribution of the values of signals intuitively, but it is very complex to estimate the parameter. Since PDF is related to the order statistics intimately (the former can be constructed by the latter approximately [53]), it has great potential to use the order statistics to construct the sparseness measure function  $J_{\mathbf{H}}$ . In fact, fourth order statistics (*kurtosis*) has been widely used to measure sparseness since the values of *kurtosis*

are easy to interpret [51], [54], [55]. But examples in [51] show that *kurtosis* based measure will encounter some risks if the distribution of the signal is not completely sure. On the other hand, Hoyer develops a measure (named as H-measure) using the first and second order statistics [36]. It is very simple, but can not reflect the distribution of signals directly [36]. Considering that the sparseness measure function should be easy to be optimized and the values should be easy to interpret, We propose a new statistics or norm based measure in this section.

Note that the Lemma in the Appendix shows some special characteristics of the function  $f = \sum_{i=1}^n (x_i^4 - \sigma_1 L^2 x_i^2 + \sigma_2 L x_i^3)$ , where  $L = \sum_{i=1}^n x_i$  reflects the L1-norm of the vector composed of  $x_1, \dots, x_n$  (for that  $\forall i, x_i > 0$ ) and  $\sigma_1 > 0, \sigma_2 = (2\sigma_1 - 4)/3$  are parameters which are used to balance the terms in  $f$ . Based on the results of the Lemma and its proof, the value of the function  $f$  interpolates smoothly between the maximum  $((1/n^3) - (\sigma_1/n) + (\sigma_2/n^2))L^4$  (it holds when all of the variables  $x_1, \dots, x_n$  are equal to  $(L/n)$ ) and the minimum  $(1 - \sigma_1 + \sigma_2)L^4$  (it holds when one of the variables is equal to  $L$  and the rest are zeros). Physically, the former denotes a signal with the smallest sparseness degree, while the latter denotes a signal with the largest sparseness degree. Motivated by this phenomenon, for the non-zero vector  $\mathbf{x} \in \Re^n$  ( $n \geq 2$ ), a new sparseness measure (S-measure) is proposed as follows:

$$S(\mathbf{x}) = \frac{f_{\max} - (k_4 - \sigma_1 k_1^2 k_2 + \sigma_2 k_1 k_3)}{f_{\max} - f_{\min}} \quad (4)$$

where  $\sigma_1 > 0, \sigma_2 = (2\sigma_1 - 4)/3$  are the bounded constants,  $k_1 = \|\mathbf{x}\|_1, k_2 = \|\mathbf{x}\|_2^2, k_3 = \|\mathbf{x}\|_3^3, k_4 = \|\mathbf{x}\|_4^4$ ,  $\|\cdot\|$  denotes L-norm,  $f_{\max} = ((1/n^3) - (\sigma_1/n) + (\sigma_2/n^2))k_1^4$ , and  $f_{\min} = (1 - \sigma_1 + \sigma_2)k_1^4$ .

S-measure interpolates smoothly between 0 and 1, and larger S-measure corresponds to sparser vector (see Fig. 1). Note that in solving SU, the quantity  $k_1$  equals to 1 (due to ASC) and the elements of the abundance matrix are nonnegative. These features greatly simplify the S-measure.

#### B. NMF-SMC Model and Algorithm

From (3), (4) and the three features of  $\mathbf{W}$  and  $\mathbf{H}$ , the following model is used for solving SU:

$$\begin{aligned} \text{Min : } D &= \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_2^2 - \alpha J(\mathbf{H}) \\ \text{s.t. } w_{ij} &\geq 0, h_{jt} \geq 0, \sum_j h_{jt} = k_1 = 1, \\ &\forall i \in \mathcal{M}, j \in \mathcal{R}, t \in \mathcal{N} \end{aligned} \quad (5)$$

where  $J(\mathbf{H}) = (1/n) \sum_{t=1}^n S\text{-measure}(\mathbf{h}_t)$  measures the sparseness of the matrix  $\mathbf{H}$  ( $\mathbf{h}_t$  denotes  $t$ th column of  $\mathbf{H}$ ),  $\alpha > 0$ .

Using the method in [1], [23], ASC can be added to the cost function. Neglecting the constant terms, the model (5) can be changed as:

$$\begin{aligned} \text{Min : } D &= \frac{1}{2} \|\tilde{\mathbf{V}} - \tilde{\mathbf{WH}}\|_2^2 + \lambda \tilde{J}(\mathbf{H}) \\ \text{s.t. } w_{ij} &\geq 0, h_{jt} \geq 0, \forall i \in \mathcal{M}, j \in \mathcal{R}, t \in \mathcal{N} \end{aligned} \quad (6)$$

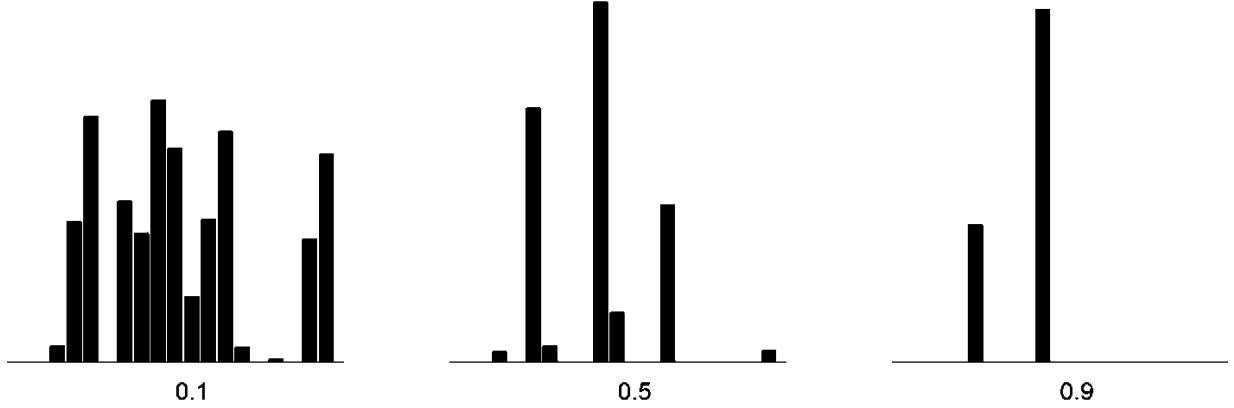


Fig. 1. Illustration of various degrees of sparseness. The sparseness degrees of these three vectors (from left to right) are 0.1, 0.5, 0.9, respectively (using S-measure).

where

$$\tilde{\mathbf{V}} = \begin{bmatrix} \mathbf{V} \\ \delta \mathbf{1}_{m \times n} \end{bmatrix}, \quad \tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W} \\ \delta \mathbf{1}_{m \times r} \end{bmatrix}, \quad \lambda, \delta > 0$$

and

$$\tilde{J}(\mathbf{H}) = \sum_{t=1}^n \sum_{j=1}^r (h_{jt}^4 - \sigma_1 h_{jt}^2 + \sigma_2 h_{jt}^3),$$

$$\sigma_2 = (2\sigma_1 - 4)/3, \quad \sigma_1 > 0.$$

Traditional ALS rule in [43] is used for solving the model (6). Let  $J_1 = (1/2)\|\tilde{\mathbf{V}} - \tilde{\mathbf{W}}\mathbf{H}\|_2^2$ ,  $J_2 = \tilde{J}(\mathbf{H})$ . The gradients of  $J_1$ ,  $J_2$  with respect to  $\tilde{\mathbf{W}}$  and  $\mathbf{H}$  are given as follows:

$$\begin{cases} \frac{\partial J_1}{\partial \tilde{\mathbf{W}}} = \tilde{\mathbf{W}}\mathbf{H}\mathbf{H}^T - \tilde{\mathbf{V}}\mathbf{H}^T \\ \frac{\partial J_1}{\partial \mathbf{H}} = \tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\mathbf{H} - \tilde{\mathbf{W}}^T\tilde{\mathbf{V}} \end{cases} \quad (7)$$

$$\begin{cases} \frac{\partial J_2}{\partial \tilde{\mathbf{W}}} = 0_{m \times r} \\ \frac{\partial J_2}{\partial \mathbf{H}} = 4\mathbf{H} \wedge 3 - 2\sigma_1 \mathbf{H} + 3\sigma_2 \mathbf{H} \wedge 2 \end{cases} \quad (8)$$

Then, the gradients of the function  $D$  in the model (6) with respect to  $\tilde{\mathbf{W}}$  and  $\mathbf{H}$  are:

$$\frac{\partial D}{\partial \tilde{\mathbf{W}}} = \tilde{\mathbf{W}}\mathbf{H}\mathbf{H}^T - \tilde{\mathbf{V}}\mathbf{H}^T \quad (9)$$

$$\begin{aligned} \frac{\partial D}{\partial \mathbf{H}} &= \frac{\partial J_1}{\partial \mathbf{H}} + \lambda \frac{\partial J_2}{\partial \mathbf{H}} \\ &= \tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\mathbf{H} - \tilde{\mathbf{W}}^T\tilde{\mathbf{V}} \\ &\quad + \lambda(4\mathbf{H} \wedge 3 - 2\sigma_1 \mathbf{H} + 3\sigma_2 \mathbf{H} \wedge 2) \end{aligned} \quad (10)$$

The gradient based updating rule for  $\tilde{\mathbf{W}}$  and  $\mathbf{H}$  is:

$$\tilde{\mathbf{W}} = \tilde{\mathbf{W}} - \eta_{\tilde{\mathbf{W}}} \frac{\partial D}{\partial \tilde{\mathbf{W}}} = \tilde{\mathbf{W}} - \eta_{\tilde{\mathbf{W}}} (\tilde{\mathbf{W}}\mathbf{H}\mathbf{H}^T - \tilde{\mathbf{V}}\mathbf{H}^T) \quad (11)$$

$$\begin{aligned} \mathbf{H} &= \mathbf{H} - \eta_{\mathbf{H}} \frac{\partial D}{\partial \mathbf{H}} = \mathbf{H} - \eta_{\mathbf{H}} (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\mathbf{H} - \tilde{\mathbf{W}}^T\tilde{\mathbf{V}}) \\ &\quad - \eta_{\mathbf{H}} \lambda(4\mathbf{H} \wedge 3 - 2\sigma_1 \mathbf{H} + 3\sigma_2 \mathbf{H} \wedge 2) \end{aligned} \quad (12)$$

To constrain  $\tilde{\mathbf{W}}$  and  $\mathbf{H}$  to be nonnegative at each iteration, let  $\sigma_1 \geq 2$  (such that  $\sigma_2 \geq 0$ ), and let  $\eta_{\tilde{\mathbf{W}}} = \tilde{\mathbf{W}} \oslash (\tilde{\mathbf{W}}\mathbf{H}\mathbf{H}^T + \beta)$ ,  $\eta_{\mathbf{H}} = \mathbf{H} \oslash (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\mathbf{H} + \beta)$ , then (11) and (12) can be simplified as:

$$\tilde{\mathbf{W}} = \tilde{\mathbf{W}} \odot (\tilde{\mathbf{V}}\mathbf{H}^T) \oslash (\tilde{\mathbf{W}}\mathbf{H}\mathbf{H}^T + \beta) \quad (13)$$

$$\begin{aligned} \mathbf{H} &= \mathbf{H} \odot (\tilde{\mathbf{W}}^T\tilde{\mathbf{V}} + 2\lambda\sigma_1 \mathbf{H}) \\ &\quad \oslash (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\mathbf{H} + \lambda(4\mathbf{H} \wedge 3 + 3\sigma_2 \mathbf{H} \wedge 2) + \beta) \end{aligned} \quad (14)$$

where  $\odot$  and  $\oslash$  denote component-wise multiplication and division [56], [57];  $\wedge$  denotes component-wise power, and  $\beta$  is a small positive constant to avoid the numerical instabilities (it is often the square root of the machine precision) [56].

According to the analysis above, the structure of the NMF-SMC algorithm is:

- 1) Preprocessing: delete the samples which are significantly influenced by noises from the collected data, then set parameter  $\delta$  and construct  $\tilde{\mathbf{V}}$  based on the rest samples  $\mathbf{V}$ ;
- 2) Initialization: let the initial values of  $\mathbf{W}$ ,  $\mathbf{H}$  be nonnegative and columns of  $\mathbf{H}$  be sum-to-one, then construct  $\tilde{\mathbf{W}}$ , combining with the parameter  $\delta$  in 1);
- 3) Updating: firstly, set parameters  $\beta$ ,  $\lambda$  and  $\sigma_1 \geq 2$  ( $\sigma_2 = (2\sigma_1 - 4)/3$ ), then, update  $\tilde{\mathbf{W}}$ ,  $\mathbf{H}$  using (13) and (14) respectively. At last, let all of the entries of the last row of  $\tilde{\mathbf{W}}$  be  $\delta$ ;
- 4) Stopping: if a stopping criterion is satisfied, the algorithm stops; otherwise, goes to step 3).

Generally, we set  $\beta = 1e-9$ ,  $\lambda \in [0, 0.1]$ ,  $\delta$  is the mean value of  $\mathbf{V}$  (see details in Section IV.A) in the above algorithm. And the maximal iteration number is often used as a stopping criterion.

### C. Algorithm Analysis

NMF-SMC is a gradient-based ALS multiplication updating algorithm, where the selection of the learning rate is adaptive. The algorithm is often initialized randomly. To speed up it, one can preprocess the observations using the singular value decomposition (SVD) or VCA, and then let the preprocessing results be the initial values [1], [44].

Under the ALS updating rule, the convergence of the cost function  $D$  in the model (6) is similar to the objective function

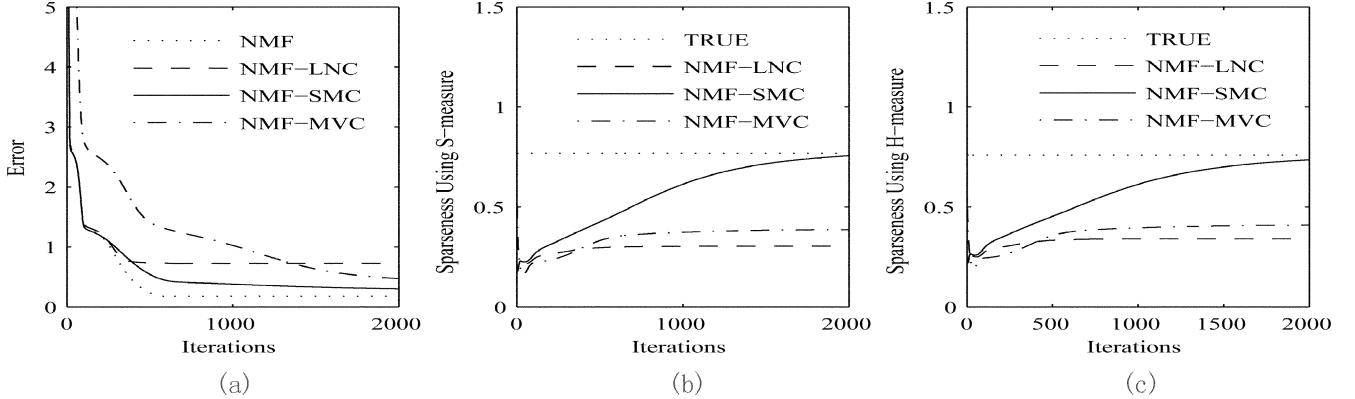


Fig. 2. The errors of the algorithms NMF, NMF-SMC, NMF-LNC, NMF-MVC at each iteration and the corresponding sparseness degree of the abundance matrix (using S-measure and H-measure, respectively) in the experiments of synthetic data (SNR = 25 dB). (a) Errors; (b) Sparseness degree using S-measure; (c) Sparseness degree using H-measure.

which is monotonically non-increasing in [43], when the parameter  $\lambda$  is a small constant or decreases with the increase of the iterations. Therefore, the proposed function  $D$  is also monotonically non-increasing if  $\lambda$  is selected properly.

From the implementation point of view, the computational load of NMF-SMC is mainly costed in the ALS updating process. For ALS approach, it is found that the computational complexity is  $O(m \times r \times n)$  approximately [58]. Therefore, the computational complexity of the proposed method is  $O(\#iter \times m \times r \times n)$  roughly, where  $\#iter$  denotes the iteration number.

#### IV. SIMULATIONS AND EXPERIMENTS

In this section, a series of experiments are carried out to show the advantages of NMF-SMC in the endmember extraction and abundance estimation. The following indices are often adopted to evaluate the algorithms: spectral angle distance (SAD), abundance angle distance (AAD) and spectral correlation (Corr) [1]. Their definitions are given, respectively, as follows:

$$\text{SAD} = \frac{180}{\pi} \cos^{-1} \left( \frac{\mathbf{w}^T \hat{\mathbf{w}}}{\|\mathbf{w}\|_2 \|\hat{\mathbf{w}}\|_2} \right) \quad (15)$$

$$\text{AAD} = \frac{180}{\pi} \cos^{-1} \left( \frac{\mathbf{h}^T \hat{\mathbf{h}}}{\|\mathbf{h}\|_2 \|\hat{\mathbf{h}}\|_2} \right) \quad (16)$$

$$\text{Corr} = \frac{(\mathbf{w} - \bar{\mathbf{w}})^T (\hat{\mathbf{w}} - \bar{\hat{\mathbf{w}}})}{\|\mathbf{w} - \bar{\mathbf{w}}\|_2 \|\hat{\mathbf{w}} - \bar{\hat{\mathbf{w}}}\|_2} \quad (17)$$

where  $\mathbf{w}$  (column vector) and  $\mathbf{h}$  (row vector) denote the real endmember signature and abundance, respectively.  $\hat{\mathbf{w}}$  and  $\hat{\mathbf{h}}$  denote the estimated endmember signature and abundance, respectively.  $\bar{\mathbf{w}}$  and  $\bar{\hat{\mathbf{w}}}$  denote the means of  $\mathbf{w}$  and  $\hat{\mathbf{w}}$ , respectively.

##### A. Synthetic Data

NMF-SMC is used to test a group of synthetic data in this simulation, where the endmember signature matrix  $\mathbf{W}_{20 \times 4}$  is selected from the USGS digital spectral library [59], and the abundance matrix  $\mathbf{H}_{4 \times 2500}$  is generated randomly using the “rand()” function in Matlab software. To simulate the possible

errors and sensor noises, some Gaussian noises are added to the mixtures, where the signal-to-noise (SNR) is calculated as:  $\text{SNR} = 10 \log(E(\mathbf{x}\mathbf{x}^T)/E(\mathbf{y}\mathbf{y}^T))$ , here  $\mathbf{x}$  denotes the real signals,  $\mathbf{y}$  denotes the corresponding noise, and  $E(\cdot)$  is the expectation operator.

First, the estimation errors and the sparseness of the estimated abundance matrices by the NMF-type algorithms NMF-SMC, NMF, NMF-LNC, and NMF-MVC are compared. As the scale of the tested dataset is small, the initial values of  $\mathbf{W}, \mathbf{H}$  are generated randomly with the same method for simplicity. For comparing with these algorithms, we have made many experiments for different noise levels. But because of the space limitation, here we can display the details for the results under  $\text{SNR} = 25$  dB, as an example. Fig. 2(a) shows the estimation errors (defined as:  $\text{Error} = \|\hat{\mathbf{V}} - \mathbf{WH}\|_F^2/2$ ) of these four algorithms, which can reflect the overall estimation precision in some sense. We can see that the comprehensive performance of the proposed NMF-SMC is the best, as far as the estimation error and the convergence speed are concerned. Fig. 2(b) shows the sparseness degree of the abundance measured by S-measure, where “True” denotes the real sparseness degree. We also compare the results with the widely used H-measure (see Fig. 2(c)). It is easy to see that the proposed measure which is with more physical meaning has similar performance to H-measure (as the sparseness curves of NMF and NMF-LNC are nearly overlapping, only the latter retains in Fig. 2(b) and (c)), and the retrieved sparseness degree of the abundance is also closer to real one.

Then, the sensitivities of NMF-SMC to the noises are compared with NMF-LNC, NMF-MVC and SC-FCLS which is composed of the endmember extraction algorithm NMF-SC in [27] and the abundance estimation algorithm FCLS [23]. The averaged AAD, SAD and Corr (termed as AVG-AAD, AVG-SAD, and AVG-Corr, respectively) are utilized to evaluate the performance. For better visual comparisons, the actual AVG-SAD and AVG-AAD are transformed by:  $y = \log(1 + x)$ , where  $x$  denotes the actual value, and  $y$  denotes the transformed result [1]. This is for suppressing the big differences between the results of NMF-LNC and NMF-SMC. Fig. 3 shows the performances of the four algorithms under different noise levels ( $\text{SNR} = 12$  dB, 16 dB, 20 dB, 24 dB, 28 dB, 32 dB, 36 dB, and 40 dB). Based on the indices AVG-AAD, AVG-SAD, and

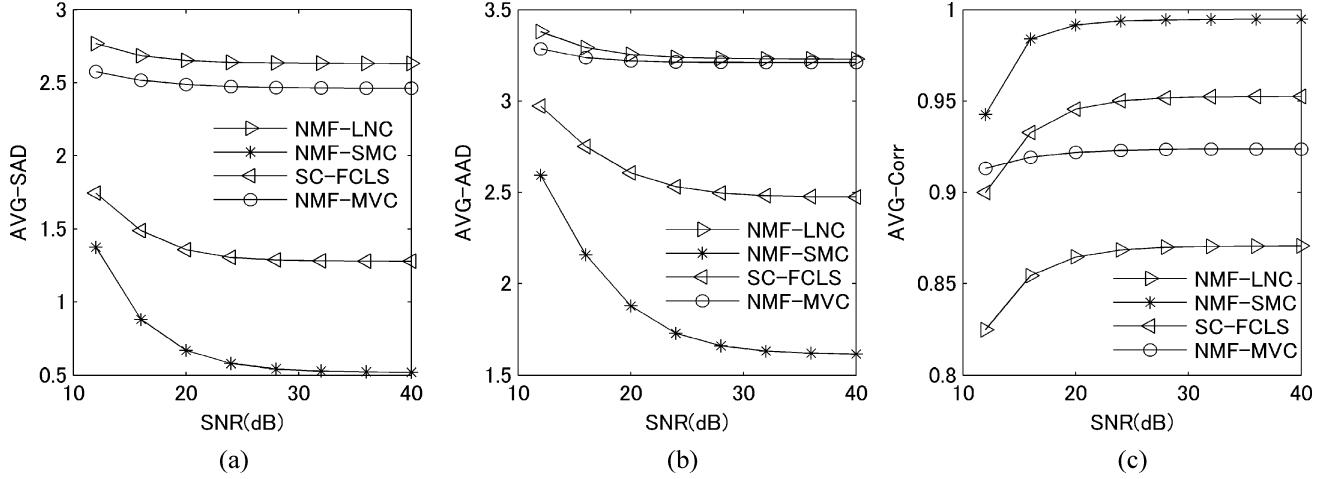


Fig. 3. Indices of AVG-SAD, AVG-AAD and AVG-Corr corresponding to the four algorithms (NMF-SMC, NMF-LNC, NMF-MVC and SC-FCLS) under the eight noise degrees (SNR = 12 dB, 16 dB, 20 dB, 24 dB, 28 dB, 32 dB, 36 dB, and 40 dB) in the experiments of synthetic data. (a) AVG-SAD; (b) AVG-AAD; (c) AVG-Corr.

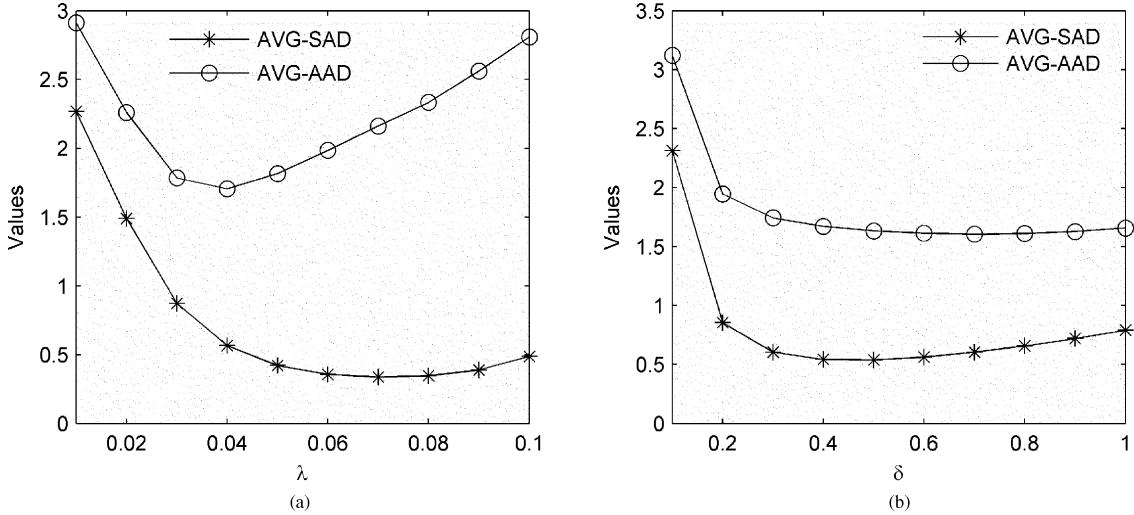


Fig. 4. Indices of AVG-SAD and AVG-AAD of the algorithm NMF-SMC for different values of parameters  $\lambda$  and  $\delta$  in the experiments of synthetic data (SNR = 25 dB). (a) AVG-SAD and AVG-AAD for different value of parameter  $\lambda$ ; (b) AVG-SAD and AVG-AAD for different value of parameter  $\delta$ .

AVG-Corr, the proposed NMF-SMC performs the best under different SNRs.

Furthermore, the sensitivity of NMF-SMC to its parameters is analyzed (for better visual comparisons, the same transformation  $y = \log(1 + x)$  as above is used for AVG-SAD and AVG-AAD). The parameters in NMF-SMC are as follows:  $\sigma_1, \sigma_2, \beta, \lambda, \delta$ , where  $\sigma_1 \geq 2$  is used to balance the norms and  $\sigma_2 = (2\sigma_1 - 4)/3$ . For the convenience of computation, it is often selected as:  $\sigma_1 = 2$  (then  $\sigma_2 = 0$ );  $\beta$  is a small positive constant to avoid numerical instabilities (generally  $\beta = 1e - 9$ ) [56];  $\lambda \in [0, 0.1]$  is a small valued parameter to balance the decomposition error and sparseness of the factor  $\mathbf{H}$ , and Fig. 4(a) shows the AVG-SAD and AVG-AAD indices using NMF-SMC under different  $\lambda$  (SNR = 25 dB);  $\delta > 0$ , controlling ASC, is used in many spectral unmixing algorithms [1], [23]. In order to balance the precision and speed of the algorithm, we select  $\delta$  to be the mean of  $\mathbf{V}$  (i.e.,  $\delta = \sum_{i=1}^n \sum_{t=1}^m v_{it}/mn$ ). Fig. 4(b) shows AVG-SAD and AVG-AAD using NMF-SMC under different  $\delta$  (SNR = 25 dB). From Fig. 4(b), the proposed

algorithm performs quite well when  $\delta$  is close to the mean 0.345 of  $\mathbf{V}$ .

Furthermore, the sensitivity of NMF-SMC to the initialization criteria is tested. Three experiments by using the criteria of randomization, non-negative double SVD (NNDSVD) in [44], and VCA in [17] are performed, respectively, where  $\sigma_1 = 2, \sigma_2 = 0, \beta = 1e - 9, \lambda = 0.04, \delta = 0.345$  (SNR = 25 dB). The corresponding results are as following: AVG-AAD = [2.2406, 2.0379, 1.8417], AVG-SAD = [0.9874, 1.1447, 0.4780], respectively. From these numerical results, one can see that the performances of the proposed NMF-SMC vary in a small range under different initialization criteria, and it tends to produce better results if VCA-based method is used.

#### B. Real Data1

In this experiment, NMF-SMC is used to test the real dataset collected by airborne visible/infrared imaging spectrometer (AVIRIS) sensor over Cuprite, Nevada [1], [17], [60]. The

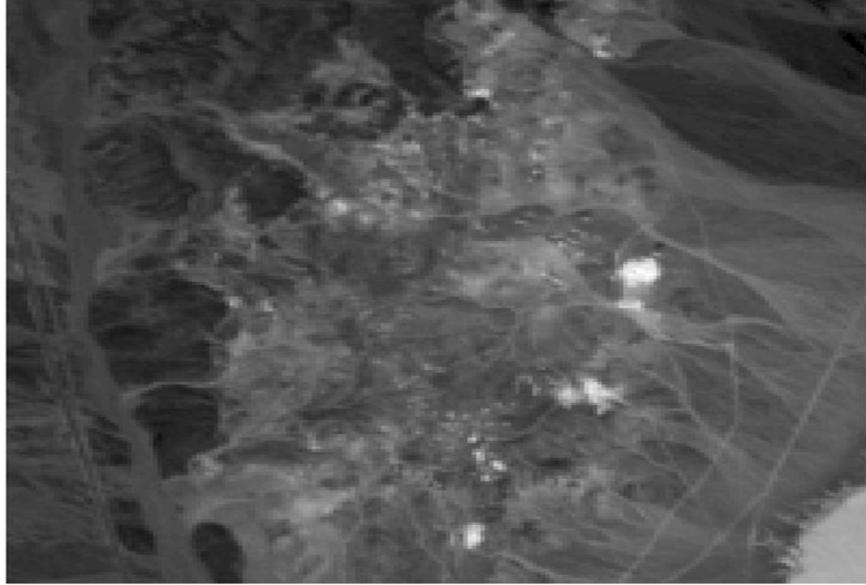


Fig. 5. Band 30 of a subimage of the AVIRIS Cuprite Nevada dataset.

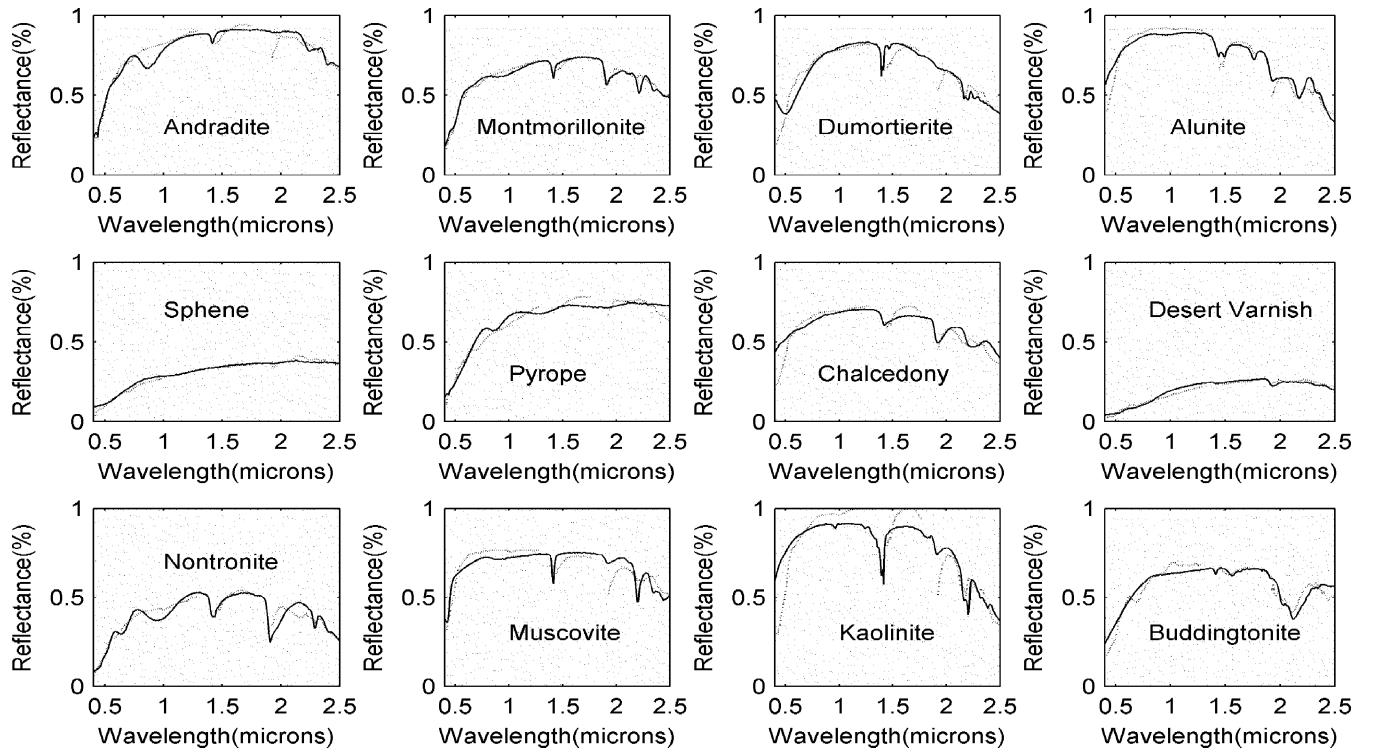


Fig. 6. Comparison of (dotted line) the extracted endmember signatures using NMF-SMC with (solid line) the USGS spectral library.

Cuprite area is a relatively undisturbed hydrothermal system with many well-exposed minerals, thus the collected dataset is suitable for remote-sensing experiments. In fact, it has been widely tested and many existing results can be used for comparison [1], [17], [24], [40], [61]. The selected sub-scene is shown in Fig. 5 (at band 30), which consists of 200 lines and 200 pixels per line. The corresponding data are collected on the AVIRIS flight June 19, 1997 [60]. Before testing NMF-SMC, the initial data is formatted into MAT by the software MultiSpec [62]. Also, in order to improve the precision, bands with low

SNR and water-vapor absorption are removed from the original dataset (including 1–3, 104–113, 148–167, 221–224). Finally, 187 bands are used in the experiment (i.e.,  $m = 187$ ).

For the experimental area, the estimated number of the endmembers is equal to 12 (i.e.,  $r = 12$ ) using the virtual dimension (VD) method in [63]. Just like NMF-MVC, we also use the results of VCA as the starting points for the NMF-SMC learning. Fig. 6 shows the comparison of the extracted endmember signatures given by NMF-SMC and those from USGS digital spectral library [59]. Based on Fig. 6,

TABLE I

SADs BETWEEN SPECTRUM FROM USGS SPECTRAL LIBRARY AND EXTRACTED SPECTRUM USING NMF-SMC, VCA AND NMF-MVC, RESPECTIVELY

	NMF-SMC	VCA	NMF-MVC
Andradite	2.8800	3.6512	5.2156
Montmorillonite	2.2664	3.2513	8.0205
Dumortierite	6.0420	3.6128	6.0245
Alunite	4.0253	5.2504	3.7217
Sphene	2.9453	3.0411	3.4203
Pyrope	3.6319	3.8526	3.9412
Chalcedony	5.8598	9.4921	6.7904
Desert Varnish	4.8858	15.3425	7.0214
Nontronite	4.8861	4.0810	6.5234
Muscovite	4.5181	5.1314	3.1012
Kaolinite	7.9032	5.7079	7.4634
Buddingtonite	4.5207	4.3247	4.2145
Mean	4.5304	5.5616	5.5382

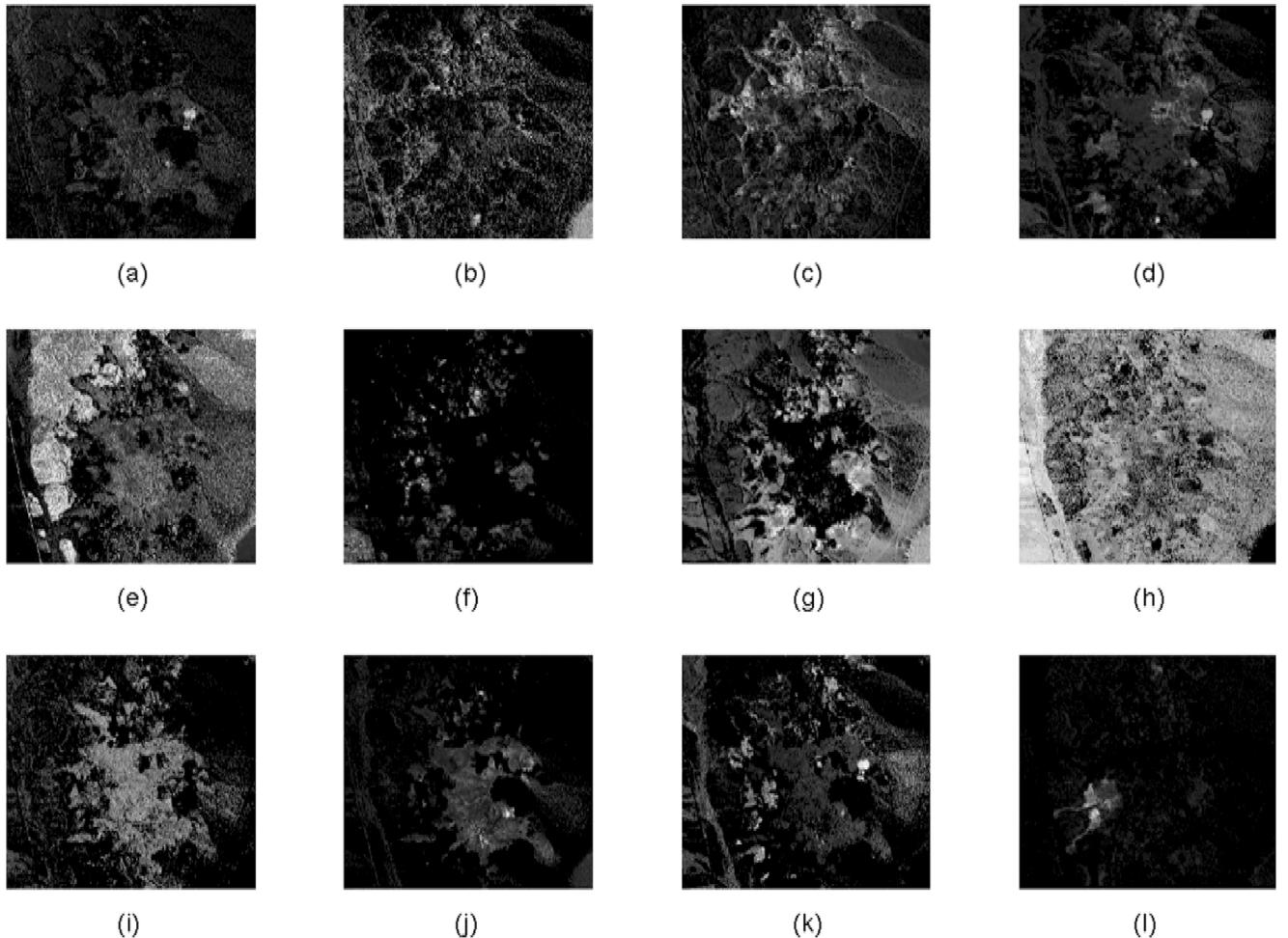


Fig. 7. The abundance maps corresponding to the extracted endmembers using NMF-SMC. (a) Andradite; (b) Montmorillonite; (c) Dumortierite; (d) Alunite; (e) Sphene; (f) Pyrope; (g) Chalcedony+Alunite; (h) Desert Varnish; (i) Nontronite; (j) Muscovite+Chalcedony+Kaolinite; (k) Kaolinite; (l) Buddingtonite.

the estimated spectrum using NMF-SMC match well with which from USGS digital spectral library. The indices SADs (calculated using estimated spectrum with which from USGS digital spectral library) of the three algorithms NMF-SMC, NMF-MVC and VCA are shown in Table I, where Mean denotes the averaged indices. Based on these results, the averaged

performance of NMF-SMC is better than NMF-MVC and VCA for endmember extraction.

Also, we analyze the corresponding abundance maps (mimicking the alteration maps shown by Ashley and Abrams [64]). Fig. 7 shows the overall abundances maps corresponding to the referred 12 endmembers, where (and throughout this Sec-

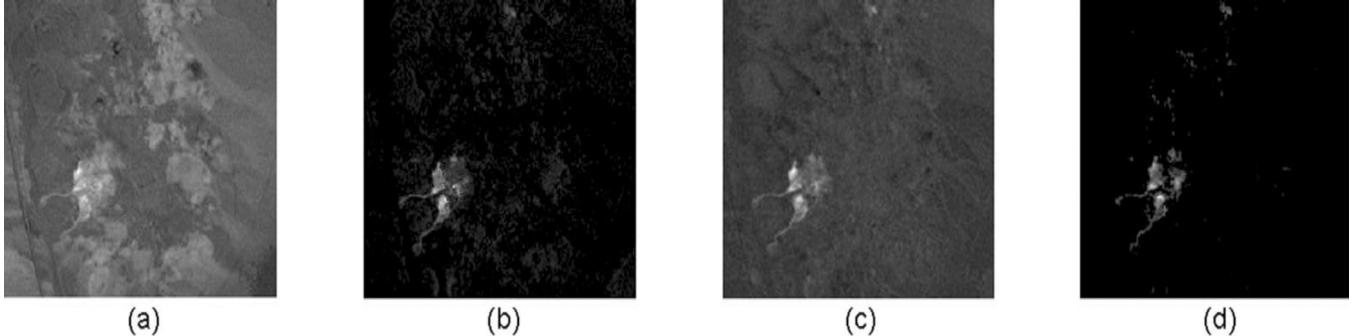


Fig. 8. Real map of Buddingtonite and the corresponding estimations of NMF-SMC, VCA-FCLS, and NMF-MVC. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) estimation of NMF-MVC.

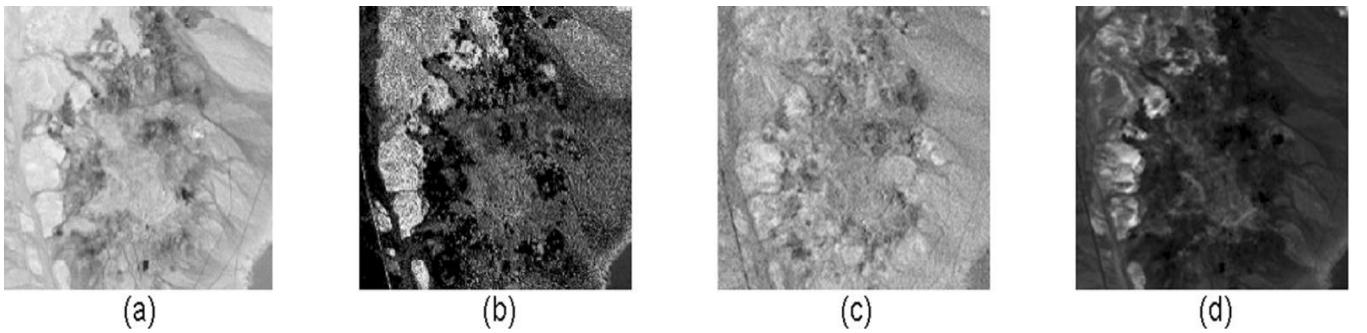


Fig. 9. Real map of Sphene and the corresponding estimations of NMF-SMC, VCA-FCLS, and GDME. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) estimation of GDME.

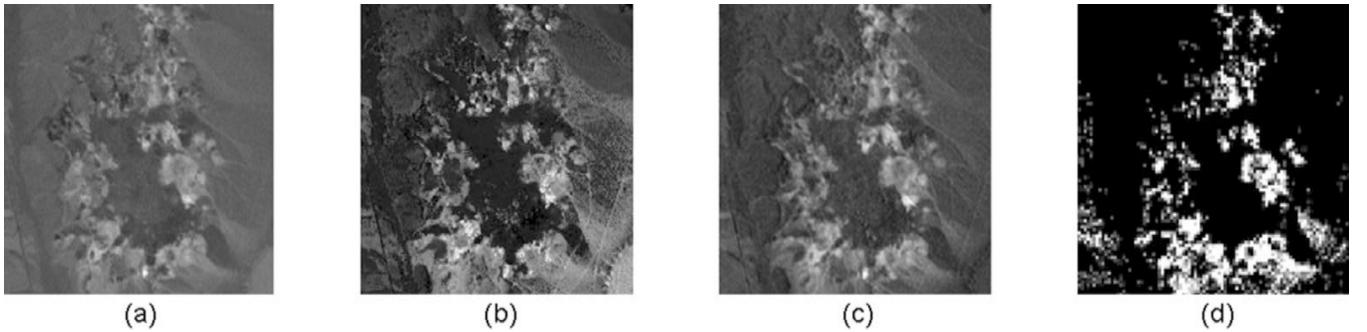


Fig. 10. Real map of Alunite, the corresponding estimations of NMF-SMC, VCA-FCLS, and Kruse's result. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) Kruse's result.

tion) pure black denotes that the percentage of a certain sort of object in this pixel is 0, while pure white denotes 1. As a whole, NMF-SMC recovers the abundances efficiently, especially for endmembers with sparse distribution (e.g., Montmorillonite, Buddingtonite, etc), although the retrieved maps about Chalcedony and Kaolinite seem not as good as other maps (the reason may be that these two endmembers interfere with each other too seriously in the dataset [1], [17]).

Moreover, we compare the retrieved maps with the existing results. As the endmember number is a little large, we mainly compare the abundance maps of three representative end-members, i.e., Buddingtonite with a small abundance fraction, Sphene with a large abundance fraction plus concentrative distribution, and Alunite with also a large abundance fraction but dispersed distribution. In order to compare with more results, for each endmember, the retrieved map is compared with the result of a different method, besides the real map obtained using library spectrum by FCLS method and the estimation of the widely used VCA-FCLS. Figs. 8–10 show abundance



Fig. 11. Band 80 of a subimage of the HYDICE urban dataset.

TABLE II  
SADs BETWEEN LIBRARY SPECTRUM AND EXTRACTED SPECTRUM USING NMF-SMC, VCA AND SCBSS, RESPECTIVELY

	NMF-SMC	VCA	SCBSS
Asphalt	11.5451	16.9023	26.3331
Roof	13.4702	29.3125	10.1471
Grass	17.2518	13.2640	12.4905
Tree	4.9045	5.7582	10.4851
Mean	11.7929	16.3092	14.8640



Fig. 12. Real map of Asphalt and the corresponding estimations of NMF-SMC, VCA-FCLS, and SCBSS. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) estimation of SCBSS.



Fig. 13. Real map of Roof and the corresponding estimations of NMF-SMC, VCA-FCLS, and SCBSS. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) estimation of SCBSS.

maps about these three endmembers, respectively. Based on the visual comparison, the proposed NMF-SMC is better than GDME in [24]. The possible reason is that the result of the gradient based GDME is affected by the learning rate which is not selected adaptively. It also shows that NMF-SMC is comparable to VCA-FCLS in [17], NMF-MVC in [1], and Kruse's result in [61] for abundance estimation.

### C. Real Data2

In this experiment, NMF-SMC is used to test the real dataset collected by Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor over urban image scene with the size  $307 \times 307$  (see Fig. 11) [65]. This dataset is composed of 210 spectral channels with a spectral resolution of 10 nm that is acquired in the 400 and 2500 nm regions. In order to improve the precision, bands with low SNR are removed from the original dataset (including channels 1–4, 76, 87, 101–111, 136–153, and 198–210), only 162 bands are used in the experiment (i.e.,  $m = 162$ ) [25]. The tested region contains a mixture of man-made objects and forestry, with four targets of interest: Asphalt, Roof, Grass, and Tree. Again, the result of VCA is used as the starting points for the NMF-SMC learning.

In order to verify the usefulness of the proposed method, we first compare the precision of the estimated endmembers with the latest published results by VCA in [17], SCBSS in [25]. Table II shows the corresponding SAD indices (calculated using estimated spectrum with library spectrum), where Mean denotes the averaged indices. It denotes that NMF-SMC performs the best. Then, we compare the retrieved maps with which of VCA-FCLS and SCBSS, plus the real maps obtained using library spectrum by FCLS method [25], [65]. Figs. 12–15 show the real maps of Asphalt, Roof, Grass, and Tree, respectively, plus the corresponding estimations using NMF-SMC, VCA-FCLS, and SCBSS. By visual comparison, the results of NMF-SMC are better than VCA-FCLS and comparable to SCBSS. Although the combination of the spectral complexity and SCBSS (i.e., SS-CBSS [25]) can improve the precision of SCBSS, there are some negative values which are obvious meaningless in the results of both SCBSS and SS-CBSS.

Considering the precision of the estimated endmember, the visualization of the maps, and the physical meaning of the estimated results, as a whole, the proposed method has the best integrated performance among the mentioned methods. Note that the target Tree is prevalent relatively but the corresponding estimated map by NMF-SMC is closer to the real map. It shows



Fig. 14. Real map of Grass and the corresponding estimations of NMF-SMC, VCA-FCLS, and SCBSS. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) estimation of SCBSS.



Fig. 15. Real map of Tree and the corresponding estimations of NMF-SMC, VCA-FCLS, and SCBSS. (a) real map; (b) estimation of NMF-SMC; (c) estimation of VCA-FCLS; (d) estimation of SCBSS.

that the proposed method is robust in the sense that it can reasonably estimate the abundance of the endmember which is not sufficiently sparse.

## V. CONCLUSION

In this paper, a novel method NMF-SMC is proposed for solving SU model blindly, where a new sparseness measure (named as S-measure) that features physical significance is developed. It is a gradient based algorithm and the learning rate is selected adaptively. As the multiplication updating rule is utilized, the results can always keep the meaningful nonnegativity. Also, there is no pure pixel assumption in NMF-SMC and there is no need for the preprocessing of dimension reduction in which some useful information may be lost. Yet, unlike the sparseness based PSNMFSC, it does not require the prior knowledge about the exact sparseness degree of the abundances.

Several experiments by using both synthetic data and real hyperspectral images are performed to evaluate the proposed NMF-SMC. Based on the results of synthetic dataset, NMF-SMC obtains results with higher precision and with faster convergence speed than that of NMF-LNC and NMF-MVC. It is also more robust to the noises than NMF-LNC, NMF-MVC, and SC-FCLS. From the results of real dataset, as a whole, NMF-SMC performs better than many published methods, including GDME, VCA-FCLS, NMF-MVC, and SCBSS. It is particularly suitable for processing sparse endmembers, but also robust to process the endmember that is not sufficiently sparse.

However, as NMF-SMC is a gradient based algorithm, the convergence speed may be slow for large scale dataset (just like NMF-MVC). Yet, if the spatial resolution of the sensors is too low, the estimation precision may be affected because the sparseness degree of the abundances may be quite small. In future work, more efficient learning algorithms for processing large dataset and low resolution images will be investigated.

## APPENDIX

*Lemma:* For  $n \geq 2, \forall i \in \mathcal{N}, x_i \geq 0$ , let  $L = \sum_{i=1}^n x_i, \phi_1 = \{(x_1, \dots, x_n) | \forall i \in \mathcal{N}, x_i = L/n\}$ , and  $\phi_2 = \{(x_1, \dots, x_n) | \exists i \in \mathcal{N}, x_i = L\}$ , then the equilibrium points set of function  $f = \sum_{i=1}^n (x_i^4 - \sigma_1 L^2 x_i^2 + \sigma_2 L x_i^3)$  is  $\phi_1 \cup \phi_2$ , where  $\sigma_1 > 0, \sigma_2 = (2\sigma_1 - 4)/3$ . Moreover, points in  $\phi_1$  are global maxima and points in  $\phi_2$  are global minima.

*Proof:*

- 1) Proof of the equilibrium points set of the function  $f$  (denote  $\mathcal{N}_0 = \{1, \dots, n-1\}\}:$

Let  $g_i = x_i^4 - \sigma_1 L^2 x_i^2 + \sigma_2 L x_i^3$ . As  $\sum_{i=1}^n x_i = L, x_n = L - \sum_{i=1}^{n-1} x_i$ . Then,  $f$  can be rewritten as:

$$f = \sum_{i=1}^{n-1} g_i + g_n \quad (18)$$

In (18),  $f$  is a function respect to  $x_1, \dots, x_{n-1}$ . If  $(x_1, \dots, x_{n-1}, L - \sum_{i=1}^{n-1} x_i)$  is the equilibrium point of  $f$ , it must hold that

$$\frac{\partial f}{\partial x_i} = 0, \quad \forall i \in \mathcal{N}_0 \quad (19)$$

On the one hand, when  $n > 2$ , from (18), for  $\forall i, j \in \mathcal{N}_0$ , it holds that

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \frac{\partial g_i}{\partial x_i} + \frac{\partial g_n}{\partial x_n} \frac{\partial x_n}{\partial x_i} \\ &= 4x_i^3 + 3\sigma_2 L x_i^2 - 2\sigma_1 L^2 x_i \\ &\quad - (4x_n^3 + 3\sigma_2 L x_n^2 - 2\sigma_1 L^2 x_n) \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{\partial f}{\partial x_j} &= \frac{\partial g_j}{\partial x_j} + \frac{\partial g_n}{\partial x_n} \frac{\partial x_n}{\partial x_j} \\ &= 4x_j^3 + 3\sigma_2 L x_j^2 - 2\sigma_1 L^2 x_j \\ &\quad - (4x_n^3 + 3\sigma_2 L x_n^2 - 2\sigma_1 L^2 x_n) \end{aligned} \quad (21)$$

Combining with (19), (20) and (21), we can get that

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \frac{\partial f}{\partial x_j} = 0 \\ &\Rightarrow 4x_i^3 + 3\sigma_2 Lx_i^2 - 2\sigma_1 L^2 x_i \\ &\quad - (4x_n^3 + 3\sigma_2 Lx_n^2 - 2\sigma_1 L^2 x_n) \\ &= 4x_j^3 + 3\sigma_2 Lx_j^2 - 2\sigma_1 L^2 x_j \\ &\quad - (4x_n^3 + 3\sigma_2 Lx_n^2 - 2\sigma_1 L^2 x_n) \\ &\Rightarrow 4x_i^3 + 3\sigma_2 Lx_i^2 - 2\sigma_1 L^2 x_i \\ &= 4x_j^3 + 3\sigma_2 Lx_j^2 - 2\sigma_1 L^2 x_j \\ &\Rightarrow 4(x_i^3 - x_j^3) + 3\sigma_2 L(x_i^2 - x_j^2) \\ &\quad - 2\sigma_1 L^2(x_i - x_j) = 0 \end{aligned}$$

Then

$$(x_i - x_j)[4(x_i^2 + x_j^2 + x_i x_j) + 3\sigma_2 L(x_i + x_j) - 2\sigma_1 L^2] = 0 \quad (22)$$

Substitute  $\sigma_2 = (2\sigma_1 - 4)/3$  to (22), it holds that

$$(x_i - x_j)[(x_i + x_j + \sigma_1 L/2)(x_i + x_j - L) - x_i x_j] = 0 \quad (23)$$

So,  $x_i = x_j$  or  $(x_i + x_j + \sigma_1 L/2)(x_i + x_j - L) = x_i x_j$ . For the latter, as  $x_i, x_j \geq 0, x_i + x_j \leq L, \sigma_1 > 0$ , such that  $x_i x_j = 0$  and  $x_i + x_j = L$ , i.e.,  $\begin{cases} x_i = 0 \\ x_j = L \end{cases}$  or  $\begin{cases} x_i = L \\ x_j = 0 \end{cases}$ . Therefore, the solution of (23) is:  $x_i = x_j$  or  $\begin{cases} x_i = 0 \\ x_j = L \end{cases}$  or  $\begin{cases} x_i = L \\ x_j = 0 \end{cases}$ .

As  $i, j \in \mathcal{N}_0$  are arbitrary, when  $n > 2$ , the solution of (19) is:  $\varphi_1 = \{(x_1, \dots, x_{n-1}) | \forall i, j \in \mathcal{N}_0, x_i = x_j\}$  or  $\varphi_2 = \{(x_1, \dots, x_{n-1}) | \exists i \in \mathcal{N}_0, x_i = L\}$ . For  $\varphi_1 : x_n = L - \sum_{i=1}^{n-1} x_i \Rightarrow x_n = L - (n-1)x_i$ , substitute  $x_n$  to  $(\partial f)/(\partial x_i) = 0$ , then we obtain

$$\begin{aligned} 4x_i^3 + 3\sigma_2 Lx_i^2 - 2\sigma_1 L^2 x_i &= 4(L - (n-1)x_i)^3 \\ &\quad + 3\sigma_2 L(L - (n-1)x_i)^2 - 2\sigma_1 L^2(L - (n-1)x_i) \end{aligned} \quad (24)$$

From (24),  $x_i = 0$  or  $x_i = L$  or  $x_i = L/n$ . As  $x_n = L - (n-1)x_i \geq 0$  and  $n > 2$ , the solution of (24) is:  $x_i = 0$  or  $x_i = L/n$ , then,  $x_n = L - 0 = L$  or  $x_n = L - (n-1)L/n = L/n$ ;

For  $\varphi_2$ :  $x_n = L - \sum_{i=1}^{n-1} x_i \geq 0 \Rightarrow x_n = 0$ .

On the other hand, when  $n = 2$ , then  $x_2 = L - x_1$ . So, (19) degenerates into:  $(\partial f)/(\partial x_1) = 0$ . Therefore,  $4x_1^3 + 3\sigma_2 Lx_1^2 - 2\sigma_1 L^2 x_1 = 4(L - x_1)^3 + 3\sigma_2 L(L - x_1)^2 - 2\sigma_1 L^2(L - x_1) \Rightarrow \varphi_3 = \{x_1 | x_1 = 0, L, L/2\}$ . Then,  $x_2 = L$  or  $x_2 = 0$  or  $x_2 = L/2$ .

Combining with  $\varphi_1, \varphi_2, \varphi_3$ , for  $n \geq 2, \forall i \in \mathcal{N}, x_i \geq 0, \sigma_2 = (2\sigma_1 - 4)/3, \sigma_1 > 0$ , the equilibrium points set of function  $f$  is  $\phi_1 \cup \phi_2$ , where  $\phi_1 = \{(x_1, \dots, x_n) | \forall i \in \mathcal{N}, x_i = L/n\}, \phi_2 = \{(x_1, \dots, x_n) | \exists i \in \mathcal{N}, x_i = L\}$ .

## 2) Proof of the extreme points of the function $f$ :

According to (20), (25), shown at the bottom of the page, holds that On the one hand, for points in  $\phi_1$ , (26) holds, shown at the bottom of the page. When  $n \geq 2, \sigma_1 > 0$ , it holds that  $((4-2n)/(n)\sigma_1 + (12-8n)/(n^2))L^2 < 0$ . Therefore, according to the decision rule to extreme points of multivariate function [66], points in  $\phi_1$  are local maxima of the function  $f$ . On the other hand, the boundary point set of the function  $f$  is composed of points in  $\phi_2$ .

As  $f$  is a bounded function in a closed interval, its global maximum and minimum exist, and they correspond to local extreme points or boundary points. Therefore, points in  $\phi_1$  are global maxima and points in  $\phi_2$  are global minima.

The Lemma is proved by combining with 1) and 2).

## ACKNOWLEDGMENT

The authors would like to thank Dr. G. A. Swayze for providing parts of his Ph.D. dissertation [40], Prof. D. A. Landgrebe and Mr. L. L. Biehl for providing the software MultiSpec [62], and Dr. J. M. B. Dias for supplying the code of VCA algorithm. The authors would also like to thank the associate editor and the anonymous reviewers for their valuable comments and suggestions.

$$\begin{cases} f''_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} = 12x_n^2 + 6\sigma_2 Lx_n - 2\sigma_1 L^2 \\ f''_{ii} = \frac{\partial^2 f}{\partial x_i^2} = 12x_i^2 + 6\sigma_2 Lx_i - 2\sigma_1 L^2 + 12x_n^2 + 6\sigma_2 Lx_n - 2\sigma_1 L^2 \end{cases} \quad (25)$$

$$\begin{cases} f''_{ij} = \frac{12L^2}{n^2} + \frac{6\sigma_2 L^2}{n} - 2\sigma_1 L^2 = \left(\frac{4-2n}{n}\sigma_1 + \frac{12-8n}{n^2}\right)L^2 \\ f''_{ii} = \frac{24L^2}{n^2} + \frac{12\sigma_2 L^2}{n} - 4\sigma_1 L^2 = 2\left(\frac{4-2n}{n}\sigma_1 + \frac{12-8n}{n^2}\right)L^2 \end{cases} \quad (26)$$

## REFERENCES

- [1] 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.
- [2] D. Landgrebe, "Multispectral land sensing: Where from, where to?," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 414–421, Mar. 2005.
- [3] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Lab. J.*, vol. 14, pp. 55–73, 2003.
- [4] J. Broadwater and R. Chellappa, "Hybrid detectors for subpixel targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1891–1903, Nov. 2007.
- [5] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, pp. 44–57, 2002.
- [6] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: Wiley, 2003.
- [7] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.
- [8] C. F. Caiafa, E. Salerno, A. N. Proto, and L. Fiumi, "Blind spectral unmixing by local maximization of non-gaussianity," *Signal Process.*, vol. 88, pp. 50–68, 2008.
- [9] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2706–2717, Nov. 1999.
- [10] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2679–2693, Aug. 2009.
- [11] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.
- [12] A. M. Filippi and R. Archibald, "Support vector machine-based end-member extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 771–791, Mar. 2009.
- [13] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.
- [14] M. E. Winter, "N-FINER: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry*, 1999, vol. 3753, pp. 266–275.
- [15] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summaries JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.
- [16] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summaries 4th Annu. JPL Airborne Geoscience Workshop*, 1993, vol. 1, pp. 11–14.
- [17] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, apr. 2005.
- [18] C.-I. Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, May 2000.
- [19] G. M. Foody, "Thematic mapping from remotely sensed data with neural networks: MLP, RBF, and PNN based approaches," *J. Geograph. Syst.*, vol. 3, pp. 217–232, 2001.
- [20] L. Bastin, "Comparison of fuzzy c-mean classification, linear mixture modeling and mlc probabilities as tools for unmixing coarse pixels," *Int. J. Remote Sens.*, vol. 18, pp. 3629–3648, 1997.
- [21] F. A. Kruse and J. W. Boardman, "Characterization and mapping of kimberlites and related diatremes using hyperspectral remote sensing," in *Proc. 2000 IEEE AeroSpace Conf.*, Big Sky, MT, 2000, pp. 299–304.
- [22] F. A. Kruse, J. H. Huntington, and R. O. Green, "Results from the 1995 AVIRIS geology group shoot," in *Proc. 2nd Int. Airborne Remote Sensing Conf. Exhibition: Environmental Research Institute of Michigan (ERIM)*, Ann Arbor, 1996, vol. I, pp. I-211–I-220.
- [23] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 529–545, 2001.
- [24] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1008–1021, Apr. 2007.
- [25] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.
- [26] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [27] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Lin. Alg. Appl.*, vol. 416, pp. 29–47, 2006.
- [28] M. Parente, A. Zymnis, J. Skaf, and J. Bishop, "Spectral unmixing with nonnegative matrix factorization," in *Proc. SPIE Conf. Remote Sensing for Env. Monitoring, GIS Appl. and Geology VI*, 2006, vol. 6366, pp. 1–7.
- [29] D. Heslop, T. Dobeneck, and M. Hocker, "Using non-negative matrix factorization in the "unmixing" of diffuse reflectance spectra," *Marine Geology*, vol. 241, pp. 63–78, 2007.
- [30] A. B. Hamza and D. J. Brady, "Reconstruction of reflectance spectra using robust nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3637–3642, Sep. 2006.
- [31] S. A. Robila and L. G. Maciak, "Considerations on parallelizing nonnegative matrix factorization for hyperspectral data unmixing," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 57–61, Jan. 2009.
- [32] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for nonnegative matrix factorization," *Comput. Stat. Data Anal.*, vol. 52, pp. 155–173, 2007.
- [33] Y. Yu, S. Guo, and W. Sun, "Minimum distance constrained nonnegative matrix factorization for the endmember extraction of hyperspectral images," in *Proc. SPIE Conf. Remote Sensing and GIS Data Process. and Appl.*, 2007, vol. 6790, pp. 151–159.
- [34] 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.
- [35] M. Heiler and C. Schnörr, "Learning sparse representations by nonnegative matrix factorization and sequential cone programming," *J. Mach. Learning Res.*, vol. 7, pp. 1385–1407, 2006.
- [36] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learning Res.*, vol. 5, pp. 1457–1469, 2004.
- [37] H. Laurberg, "Uniqueness of non-negative matrix factorization," in *Proc. IEEE Int. Conf. Stat. Signal Process.*, 2007, pp. 44–48.
- [38] R. M. D. Freitas, V. Haertel, and Y. E. Shimabukuro, "Linear spectral mixture model in moderate spatial resolution image data," *Boletim De Ciencias Geodeticas*, vol. 14, pp. 55–71, 2008.
- [39] T. W. Ray and B. C. Murray, "Nonlinear spectral mixing in desert vegetation," *Remote Sens. Environ.*, vol. 55, pp. 59–64, 1996.
- [40] G. Swayze, "The hydrothermal and structural history of the cuprite mining district, southwestern Nevada: An Integrated geological and geophysical approach," Ph.D. dissertation, Univ. Colorado, Boulder, CO, 1997.
- [41] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, pp. 542–552, 1994.
- [42] B. A. Pearlmutter and V. K. Potluri, "Sparse separation: principles and tricks," in *Proc. SPIE Int. Conf. ICA, Wavelets and Neural Netw.*, 2003, vol. 5102, pp. 1–4.
- [43] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," *Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 556–562, 2001.
- [44] C. Boutsidis and E. Gallopolous, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognit.*, vol. 41, pp. 1350–1362, 2008.
- [45] A. Cichocki, A. H. Phan, R. Zdunek, and L. Q. Zhang, "Flexible component analysis for sparse, smooth, nonnegative coding or representation," *Lecture Notes in Computer Science*, vol. 4984, pp. 811–820, 2008.
- [46] D. L. Donoho, "Sparse components analysis and optimal atomic decompositions," *Constructive Approx.*, vol. 17, pp. 353–382, 2001.
- [47] K. Stadthanner *et al.*, "Sparse nonnegative matrix factorization with genetic algorithms for microarray analysis," in *Proc. Int. Joint Conf. Neural Netw.*, 2007, pp. 294–299.
- [48] Y. Q. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, pp. 1193–1234, 2004.
- [49] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization," *PNAS*, vol. 100, pp. 2197–2202, 2003.
- [50] B. A. Olshausen and D. J. Field, "Emergence of simplecell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.
- [51] J. Karvanen and A. Cichocki, "Measure sparseness of noisy signals," in *Proc. Int. Symp. ICA and BSS*, Nara, Japan, 2003.

- [52] Z. S. He, S. L. Xie, and Y. L. Fu, "Sparse measure of signal," in *Proc. IEEE Int. Conf. Neural Netw. Brain.*, 2005, vol. 3, pp. 1931–1936.
- [53] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, pp. 287–314, 1994.
- [54] B. Willmore and D. J. Tolhurst, "Characterizing the sparseness of neural codes," *Network: Comput. Neural Syst.*, vol. 12, pp. 255–270, 2001.
- [55] N. Saito, B. M. Larson, and B. Benichou, "Sparsity vs. statistical independence from a best-basis viewpoint," in *Proc. SPIE Conf. Wavelet Appl. Sig. Image Process. VIII*, 2000, vol. 4119, pp. 474–486.
- [56] A. Cichocki, R. Zdunek, and S. Amari, "Nonnegative matrix and tensor factorization," *IEEE Signal Process. Mag.*, vol. 26, pp. 142–145, 2008.
- [57] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, 3rd ed. London, U.K.: Cambridge Univ. Press, 1991.
- [58] V. Monga and M. K. Mihcak, "Robust and secure image hashing via non-negative matrix factorization," *IEEE Trans. Inf. Forensics Security*, vol. 2, pp. 376–390, 2007.
- [59] R. N. Clark, G. A. Swazey, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. Geological Survey Digital Spectral Library: Version 1: 0.2 to 3.0  $\mu\text{m}$ " Washington, DC, Open File Rep. 93-592, 1993, U.S. Geol. Surv.
- [60] [Online]. Available: <http://aviris.jpl.nasa.gov/>
- [61] [Online]. Available: [http://www.hgimaging.com/PDF/Kruse\\_JPL2002\\_AVIRIS\\_Hyperion.pdf](http://www.hgimaging.com/PDF/Kruse_JPL2002_AVIRIS_Hyperion.pdf)
- [62] [Online]. Available: <http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/> [Online]. Available:
- [63] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, pp. 608–619, 2004.
- [64] R. P. Ashley and M. J. Abrams, "Alteration Mapping Using Multispectral Images, Cuprite Mining District, Esmeralda County, Nevada," United States Geological Survey Open-File Report 80-367, 1980.
- [65] [Online]. Available: <http://www.tec.army.mil/Hypercube/> [Online]. Available:
- [66] Z. Y. Li, Z. C. Lu, and H. Chen, "A convenient judgement method about extreme value of one class multivariate functions and its application," *Mathem. Practice Theory*, vol. 33, pp. 96–101, 2003.



**Zuyuan Yang** was born in Hubei Province, China, in 1982. He received the Ph.D. degree in signal and information processing from South China University of Technology, Guangzhou, China, in 2010.

He is currently a Postdoctor in the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China. His research interests include machine learning, blind signal processing, image processing, and information security.



**Guoxu Zhou** was born in Hubei Province, China, in 1977. He is currently working towards the Ph.D. degree at the Intelligent Information and Signal Processing Group, South China University of Technology, China, led by Prof. Shengli Xie.

His research interests include machine learning, intelligent information processing, and blind signal processing.



**Shengli Xie** (M'01–SM'02) was born in Hubei Province, China, in 1958. He received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, in 1992 and the Ph.D. degree in control theory and applications from South China University of Technology, Guangzhou, China, in 1997.

He is presently a Full Professor with the South China University of Technology and a vice head of the Institute of Automation and Radio Engineering. His research interests include automatic control, blind signal processing and machine learning. He is the author or coauthor of two books and more than 70 scientific papers in journals and conference proceedings.



**Shuxue Ding** (M'04) received the M.Sc. degree in physics from Dalian University of Technology, China, in 1988, and the Ph.D. degree in physics from Tokyo Institute of Technology, Japan, in 1996.

From 1989 to 1991 and from 1991 to 1992, respectively, he was an Assistant Professor and Associate Professor at Dalian University of Technology. From 1996 to 1998, he was with Fujisoft-ABC Inc., Japan, where he was involved in algorithm design for telecommunication systems. From 1998 to 2003, he was with Clarion Co., Ltd, Japan, where he engaged in researches in communication and signal processing, especially in speech recognition. From 2003 to 2005, he was a visiting faculty at the University of Aizu, Japan. He is currently an Associate Professor in the School of Computer Science and Engineering, the University of Aizu, and an Associate Research Scientist in the Brian Science Institute, the Institute of Physical and Chemical Research (RIKEN), Japan.

Dr. Ding has been engaged in research in a wide range of areas of mathematical and physical engineering, such as statistical signal processing, optimization, neural computation, bioelectromagnetism, and information sciences. In particular, he has devoted himself to blind source separation and independent component analysis, and their applications in acoustic signals and vital signs. Recently, he is also conducting research in brain-style information processing, pattern recognition, compressive sensing and sparse representation. He is also interested in speech and image processing, quantum computation, quantum and information, and the physics of information. He is a member of ASA, ACM, and IEICE.



**Jun-Mei Yang** was born in Shandong, China, in 1979. She received the M.S. degree in cybernetics from Chinese Academy of Sciences (CAS), Beijing, China, in 2005 and the Ph.D. degree in systems science at the Graduate School of Informatics, Kyoto University, Kyoto, Japan, in 2008.

She is currently a Lecturer in the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China. Her research interests include system identification, blind signal processing and adaptive noise cancellation.



**Jun Zhang** received the B.S. degree in 1987, the M.Sc. degree in 1991, and the Ph.D. degree in 2001, from Beihang University, Beijing, China.

He is currently a Professor and Dean in the School of Electronics and Information Engineering, Beihang University, Beijing, China. His current research interests are in the areas of integrated networks of air, ground, and sky communication systems, modern air traffic management, ad hoc networks, network management. He has authored two books, more than 60 referred papers, as well as more than 10 patents or pending applications.

Dr. Zhang currently serves on the editorial boards of CAAI Transactions on Intelligent Systems and two committees of international conferences.