Hyperspectral Unmixing viaL1/2 Sparsity-Constrained Nonnegative Matrix Factorization

Author

Qian, Yuntao, Jia, Sen, Zhou, Jun, Robles-Kelly, Antonio

Published

Journal Title

DOI

Copyright Statement

Downloaded from

Griffith Research Online

Hyperspectral Unmixing Via $L_{1/2}$ Sparsity-constrained Nonnegative Matrix Factorization

Yuntao Qian, *Member, IEEE*, Sen Jia, Jun Zhou, *Member, IEEE* and Antonio Robles-Kelly, *Senior Member, IEEE*

Abstract

Hyperspectral unmixing is a crucial preprocessing step for material classification and recognition. In the last decade, nonnegative matrix factorization (NMF) and its extensions have been intensively studied to unmix hyperspectral imagery and recover the material end-members. As an important constraint for NMF, sparsity has been modeled making use of the L_1 regularizer. Nonetheless, recent studies show that the solutions with L_1-norm are less sparse than those yielded by its L_0 counterpart, while solving the L_0-norm is an NP hard problem. Furthermore, the L_1 regularizer conflicts with the full additivity constraint of material abundances, hence, limiting the practical efficacy of NMF methods in hyperspectral unmixing. In this paper, we extend the NMF method by incorporating the $L_{1/2}$ sparsity constraint, which we name $L_{1/2}$-NMF. The $L_{1/2}$ regularizer not only induces sparsity, but is also an unbiased estimator. We provide an iterative estimation algorithm for $L_{1/2}$-NMF, which provides more sparse and accurate results than those delivered making use of the L_1 norm. We do this by considering the end-member additivity constraint explicitly in the optimization process. We illustrate the utility of our method on synthetic and real hyperspectral data.

Index Terms

Hyperspectral unmixing; Nonnegative matrix factorization; Sparse coding, $L_{1/2}$ regularizer

I. INTRODUCTION

Hyperspectral data is acquired by high spectral-resolution imaging sensors, containing hundreds of contiguous narrow spectral band images. Due to the low spatial resolution of the sensor, disparate substances may contribute to the spectrum for a single pixel, leading to the existence of "mixed" spectra in hyperspectral imagery. Hence, hyperspectral unmixing, which decomposes a mixed pixel into a collection of constituent spectra, or *end-members*, and their corresponding fractional abundances, is often employed to preprocess hyperspectral data [1]. Several hyperspectral unmixing methods have been proposed in recent years, which include N-FINDR [2], vertex component analysis [3], independent component analysis (ICA) [4], the minimum volume enclosing simplex algorithm [5] and flexible similarity measures [6].

Most of these methods assume a linear spectral mixture model for the unmixing process. If the number and signatures of endmembers are unknown, unmixing becomes a blind source separation (BSS) problem. This is compounded by the need to estimate the parameters of the mixing and/or filtering processes. It is impossible to uniquely estimate the original source signals and mixing matrix if no *a priori* knowledge is applied to the BSS. Various approaches have specific physical and statistical assumptions for modeling the unmixing process. For example, the assumption of source independence leads to the ICA method [4], whereas the assumption of Markov random distribution of abundance leads to the spatial structure method in [7]. Since the source signals are generally independent from one another, it can be assumed that the subcomponents of the sources are mutual independent, which leads to subband ICA [8].

From the linear algebra point of view, BSS is a constrained matrix factorization problem that has found numerous applications in feature and signal extraction [9]. For general matrix factorization problems, traditional matrix computation tools such as singular vector decomposition (SVD), QR decomposition and LU factorization can be used. However, these tools can not be directly applied to hyperspectral unmixing because two constraints have to be considered [10]. The first constraint is the nonnegativity of both spectra and their fractional abundances. This is natural as the contribution from end-members should be larger than or equal to zero. Secondly, the additivity constraint over the fractional abundances has to be considered, which guarantees the addition of the proportional contribution from the end members matches the mixed observation.

Nonnegative matrix factorization (NMF) [11], [12], which decomposes the data into two nonnegative matrices, is a natural solution to the nonnegativity constraint [13]. From the data analysis point of view, NMF is very attractive because it usually provides a part-based representation of the data, making the decomposition matrices more intuitive and interpretable [14], [15]. However, the solution space of NMF is too large without other constrains. This, added to the fact that the cost function is not convex, can make the algorithm prone to noise corruption and computationally demanding. To reduce the space of solutions, extensions of NMF including symmetric NMF, semi-NMF, non-smooth NMF, and multi-layer NMF have been proposed [14].

When NMF is applied to hyperspectral unmixing, sparsity constraints are favored since they allow exploiting the notion that most of the pixels are mixtures of only a few of the end-members in the scene [16], [17]. This

implies that a number of entries in the abundance matrix are zeros, which manifests itself as a large degree of sparsity. Regularization methods are usually utilized to define the sparsity constraint on the abundance of the endmembers. Along these lines, the L_0 regularizer accounts for the number of zero elements in an abundance matrix so as to yield the most sparse result given a cost function. However, the application of the L_0 regularizer is an NP hard optimization problem that cannot be solved in practice [18]. The L_2 regularizer, on the other hand, generates smooth but not sparse results [19]. In general, the L_1 regularizer is the most popular one for achieving sparsity of the abundance matrix [20], [21], [22], [23], [24].

Despite the widely use of L_1 regularizer, there are two limitations with it when applied to hyperspectral unmixing. Firstly, enforcing a sufficiently sparse solution is not straightforward since the penalty imposed upon the cost function is a linear one with respect to the deviations of the abundance matrix from zero. Secondly, and more severely, it conflicts with the full additivity unmixing constraint. Recall that the full additivity constraint requires that the sum over the L_1 regularization function be a constant. This is often conflicts with the sparsity constraint imposed by the L_1 regularizer which favours smaller, not constant, summations over the terms of the regularization function. Therefore, the L_1 regularization is not an ideal tool for sparsity characterization of end-member abundance matrix.

Recently, the properties of fractional $L_q (0 \le q \le 1)$ regularizers have been studied in [25], [26]. It can be shown that the $L_q (0 < q < 1)$ regularizers give more sparse solutions than its L_1 counterpart. Furthermore, the sparsity of the $L_q (1/2 \le q < 1)$ solution increases as q decreases, whereas the sparsity of the solution for $L_q (0 < q \le 1/2)$ is almost linear with respect to q. For $q = 1/2$, Xu *et al* [27] have shown that, in addition to strong sparsity, the resultant regularizer is an unbiased estimator.

In this paper, we introduce the $L_{1/2}$ regularization into NMF, which we name $L_{1/2}$-NMF, so as to characterize the sparsity of abundances. The $L_{1/2}$-NMF presented here is effected through the multiplicative update algorithm in [28] through iterative applications of a rescaled gradient descent approach so as to ensure convergence. In our approach, the full additivity constraint is considered in the parameter update process. The experiments on synthetic and real hyperspectral data demonstrate the effectiveness of the $L_{1/2}$-NMF approach for unmixing hyperspectral data. Our contribution is, hence, to introduce a novel NMF method which recovers a sparse solution to the unmixing problem making use of an optimization algorithm that guarantees stable convergence to a local minima.

The rest of the paper is organized as follows. In Section II, we introduce the linear spectral mixture model and give a brief background of the nonnegative matrix factorization used here. This section also presents our $L_{1/2}$-NMF model for unmixing. Section III derives the multiplicative update algorithm employing a rescaled gradient descent scheme. Implementation issues are discussed later on. These include the consideration of the unmixing additivity constraint. Results on synthetic and real-world data are reported in Sections IV-A and Section IV-B, respectively. Finally, Section V draws conclusions and suggestions on future work.

II. $L_{1/2}$-NMF Unmixing Model

In a hyperspectral image, each pixel describes the spectral radiance of the corresponding ground location. Due to the low spatial resolution of hyperspectral imagery, a pixel often covers several different materials. Therefore,

the spectral irradiance is normally a combined result of several materials according to their distributions and configurations. Unmixing aims at detecting the existence of the contributing materials in the region and estimating their proportions. To do so, the development of mixing/unmixing models is crucial. These models should consider the interpretation of the image formation process, be physically meaningful, statistically accurate and computationally feasible. In this section, we introduce the $L_{1/2}$-NMF mixture model used throughout the paper.

A. Linear spectral mixture model

The classical linear mixing model, which is used to represent the spectrum of a pixel of L wavelength-indexed bands in the observed scene [29] based upon K end-member abundances is given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e} \tag{1}$$

where $\mathbf{x}$ denotes a $L \times 1$ vector of observed pixel spectra in a hyperspectral image, $\mathbf{s}$ is a $K \times 1$ vector of abundance fractions for each end-member, $\mathbf{e}$ is a $L \times 1$ vector of an additive noise representing the measurement errors, and $\mathbf{A}$ is a $L \times K$ nonnegative spectral signature matrix whose columns correspond to an end-member spectrum.

Using matrix notation, the mixing model above for the N pixels in the image can be rewritten as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \tag{2}$$

where the matrices $\mathbf{X} \in \mathbb{R}_+^{L \times N}$, $\mathbf{S} \in \mathbb{R}_+^{K \times N}$ and $\mathbf{E} \in \mathbb{R}^{L \times N}$ represent, respectively, the hyperspectral data, the end-member abundances and additive noise. Note that, in general, only $\mathbf{X}$ is known in advance, while the other two matrices, $\mathbf{A}$ and $\mathbf{S}$ are our targets of computation. Moreover, from observation, we can see that the product in the first right-hand side term leads itself to a matrix factorization problem.

B. NMF with sparsity constraints

Nonnegative matrix factorization (NMF) has received considerable attention in the fields of pattern recognition and machine learning, where it leads to a "part-based" representation since it allows only additive, not subtractive, combination of factors. Linear mixing models assume that the hyperspectral image is constituted of spectral signatures of end-members with corresponding nonnegative abundances. Therefore, the non-negativity of $\mathbf{A}$ and $\mathbf{S}$ mentioned above is a natural property of the measured quantities in hyperspectral data. This non-negativity can replace the independence constraint used for BSS and exploited by methods such as ICA.

To obtain $\mathbf{A}$ and $\mathbf{S}$, NMF can be performed by minimising the difference between $\mathbf{X}$ and $\mathbf{A}\mathbf{S}$ and enforcing non-negativity on $\mathbf{A}$ and $\mathbf{S}$. Such difference is often measured making use of the Euclidean distance, relative entropy or Kullback Leibler divergence. The loss function for NMF based upon the Euclidean distance is as follows

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_2^2 \tag{3}$$

Although there are numerous optimization algorithms to estimate $\mathbf{A}$ and $\mathbf{S}$, it is difficult to obtain a globally optimal solution because of the non-convexity of $\mathcal{C}(\mathbf{A}, \mathbf{S})$ with respect to both $\mathbf{A}$ and $\mathbf{S}$. Moreover, NMF is always

utilized with other constraints, such as sparsity. This is due to the fact that NMF lacks a unique solution. This can be easily verified by considering $\mathbf{AS} = (\mathbf{AD})(\mathbf{D}^{-1}\mathbf{S})$ for any nonnegative invertible matrix $\mathbf{D}$.

Note that NMF with sparsity constraints has been used as an effective tool for dimensionality reduction, feature extraction and source separation [14]. A sparse representation of the data by a limited number of components is supported from many fields such as statistics, microeconomics, biology, artificial intelligence, and information retrieval. Studies have shown that sparse coding provides a set of spatially localized, oriented and bandpass representation, similar to those found in primary visual processing [30]. The importance of sparsity is also illustrated in compressive sensing [31]. Likewise, sparsity is an intrinsic property of hyperspectral data. In most cases, the abundance distribution of any end-member does not apply to the scene as a whole. This implies that the mixed pixel is usually the superposition of only a few end-members, not all those present in the scene. That is, for each end-member, its abundance is localized with a degree of sparseness.

Thus, here we consider NMF with a sparsity constraint as an objective function for our minimization problem. This objective function is the combination of the reconstruction error and a sparsity measure as follows

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_2^2 + \lambda f(\mathbf{S}) \tag{4}$$

where $\lambda \in \mathbb{R}^+$ is a scalar that weights the contribution of the sparsity measure function $f(\cdot)$ of the matrix $\mathbf{S}$, which is usually regarded as a regularisation term.

C. NMF with $L_{1/2}$ regularizer

Many forms of regularizers $f(y)$ exist such that sparsity is encouraged. In recent years, there has been an increasing interest in the L_1 regularizer since it yields sparse solutions that are easily interpreted with a reasonable sample complexity, i.e. the number of training samples required to recover a solution grows logarithmically with respect to the number of outliers in the set. This indicates that the L_1 regularizer can be effective with small sample-sizes in a high-dimensional space. Furthermore, the L_1 regularizer has a better asymptotic sample-consistency than its L_2 counterpart. However, for spectral unmixing, the L_1 regularizer is not consist with full additivity constraint. Finding new regularizers that yield sparse solutions while preserving the additivity constraint over the end-members is a capital problem in NMF-based unmixing methods.

Here, we explore the use of the $L_{1/2}$ regularizer is an alternative to L_1. As mentioned earlier, the $L_{1/2}$ regularizer is a sparsity-promoting function [27]. Further, the $L_{1/2}$ regularizer not only can provide sparse solutions close to those yielded when L_0 is used, but is also computationally efficient. Based on Equation (4), the $L_{1/2}$-NMF model for unmixing is given by

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_2^2 + \lambda\|\mathbf{S}\|_{1/2} \tag{5}$$

where

$$\|\mathbf{S}\|_{1/2} = \sum_{k,n=1}^{K,N} \mathbf{s}_n(k)^{1/2} \tag{6}$$

and $\mathbf{s}_n(k)$ is the abundance fraction for the kth end-member at the nth pixel in the image.

III. ALGORITHM FOR $L_{1/2}$-NMF BASED UNMIXING

With the equation above at hand, we now focus on achieving a factorization solution compliant with the additivity constraint over the end-members. Thus, in this section, we first propose a general multiplicative iterative algorithm for $L_{1/2}$-NMF and prove its convergence. Then, we discuss implementation issues related to explicitly including the additivity unmixing constraint into the optimisation process, parameter initialisation, estimation of the number of end-members and the setup of the parameter λ.

A. Multiplicative Iterative Algorithm for $L_{1/2}$-NMF

The objective function described in Equation (5) includes a quadratic error term added to a sparsity-inducing regularizer. This cost function is convex with respect to the individual parameters $\mathbf{A}$ and $\mathbf{S}$. The most popular algorithms for solving NMF are iterative ones which minimize a multi-variate objective function by dividing the parameters into two sets and adopting a dual-step process. In the first step, a subset of the parameters is updated while the other set remains fixed. The second step proceeds conversely by fixing the newly updated parameters while estimating the solution of the second subset. Despite effective, such alternating solutions have a numbers of drawbacks, such as slow and unstable convergence and susceptibility to spurious local minima.

Multiplicative iterative algorithms have relatively low complexity and overcome some of these drawbacks. Lee and Seung derived a multiplicate update rule for standard NMF whose convergence has been proved [28]. When applied to Equation (3) this multiplicative update becomes

$$\mathbf{A} \quad \leftarrow \quad \mathbf{A}. * \mathbf{X}\mathbf{S}^T./\mathbf{A}\mathbf{S}\mathbf{S}^T \tag{7}$$

$$\mathbf{S} \quad \leftarrow \quad \mathbf{S}. * \mathbf{A}^T\mathbf{X}./\mathbf{A}^T\mathbf{A}\mathbf{S} \tag{8}$$

An extension of the above multiplicative rule for L_1-NMF was later developed by Hoyer [32] as follows

$$\mathbf{A} \quad \leftarrow \quad \mathbf{A}. * \mathbf{X}\mathbf{S}^T./\mathbf{A}\mathbf{S}\mathbf{S}^T \tag{9}$$

$$\mathbf{S} \quad \leftarrow \quad \mathbf{S}. * \mathbf{A}^T\mathbf{X}./(\mathbf{A}^T\mathbf{A}\mathbf{S} + \lambda) \tag{10}$$

where $(\cdot)^T$ denotes the transpose of the matrix, .* and ./ denote element-wise multiplication and division, respectively.

As for the $L_{1/2}$-NMF in Equation (5), the rescaled gradient descent introduced in [32] can be achieved by a modification of the above multiplicative update rules as follows

$$\mathbf{A} \quad \leftarrow \quad \mathbf{A}. * \mathbf{X}\mathbf{S}^T./\mathbf{A}\mathbf{S}\mathbf{S}^T \tag{11}$$

$$\mathbf{S} \quad \leftarrow \quad \mathbf{S}. * \mathbf{A}^T\mathbf{X}./(\mathbf{A}^T\mathbf{A}\mathbf{S} + \frac{\lambda}{2}\mathbf{S}^{-\frac{1}{2}}) \tag{12}$$

Likewise, where $\mathbf{S}^{-\frac{1}{2}}$ is given by the sum over the element-wise square root for each entry in the matrix $\mathbf{S}$.

Note that the update rule for $\mathbf{A}$ in (11) is the same for all three versions of NMF. For the sake of brevity, we focus our attention on the update rule for $\mathbf{S}$ in Equation (12). To make our elaboration clearer, we focus on each column of $\mathbf{S}$ alone. We can do this without any loss of generality since the objective function is separable in the

columns of $\mathbf{S}$. For convenience, let these columns be denoted $\mathbf{s}$. Similarly, the corresponding row of $\mathbf{X}$ is denoted $\mathbf{x}$. The column-wise objective function becomes

$$\mathcal{C}(\mathbf{s}) = \frac{1}{2}\|\mathbf{x} - \mathbf{As}\|_2^2 + \lambda\|\mathbf{s}\|_{\frac{1}{2}} \tag{13}$$

To guarantee the convergence of the update rule in Equation (12), we now proceed to show that the objective function decreases monotonically. To do so, we define an auxiliary function $G(\mathbf{s}, \mathbf{s}^t)$ satisfying the conditions $G(\mathbf{s}, \mathbf{s}) = \mathcal{C}(\mathbf{s})$ and $G(\mathbf{s}, \mathbf{s}^t) \geq \mathcal{C}(\mathbf{s})$ such that $\mathcal{C}(\mathbf{s})$ is non-increasing when updated using the following equation

$$\mathbf{s}^{(t+1)} = \arg\min_{\mathbf{s}} G(\mathbf{s}, \mathbf{s}^t) \tag{14}$$

This is guaranteed by

$$\mathcal{C}(\mathbf{s}^{(t+1)}) \leq G(\mathbf{s}^{(t+1)}, \mathbf{s}^t) \leq G(\mathbf{s}^t, \mathbf{s}^t) = \mathcal{C}(\mathbf{s}^t) \tag{15}$$

Following [32], we define the function G as

$$G(\mathbf{s}, \mathbf{s}^t) = \mathcal{C}(\mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t)(\nabla\mathcal{C}(\mathbf{s}^t))^T + \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)\mathbf{K}(\mathbf{s}^t)(\mathbf{s} - \mathbf{s}^t)^T \tag{16}$$

where the diagonal matrix $\mathbf{K}(\mathbf{s}^t)$ is given by

$$\mathbf{K}(\mathbf{s}^t) = \mathrm{diag}\left(\left(\mathbf{A}^T\mathbf{As}^t + \frac{\lambda}{2}(\mathbf{s}^t)^{-\frac{1}{2}}\right)./\mathbf{s}^t\right) \tag{17}$$

Here, $\mathrm{diag}(\mathbf{s})$ denotes the matrix whose diagonal is given by the entries of the vector $\mathbf{s}$ and off diagonal elements are null. Since $G(\mathbf{s}, \mathbf{s}) = \mathcal{C}(\mathbf{s})$, the Taylor expansion of $\mathcal{C}(\mathbf{s})$ is given by

$$\mathcal{C}(\mathbf{s}) = \mathcal{C}(\mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t)(\nabla\mathcal{C}(\mathbf{s}^t))^T + \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)\left(\mathbf{A}^T\mathbf{A} - \frac{\lambda}{4}\mathrm{diag}\left((\mathbf{s}^t)^{-\frac{3}{2}}\right)\right)(\mathbf{s} - \mathbf{s}^t)^T + \mathcal{R}\left(\nabla^{(n\geq 3)}\mathcal{C}(\mathbf{s}^t)\right) \tag{18}$$

where the function $\mathcal{R}$ denotes the Lagrange remainder. Note that the constraint $G(\mathbf{s}, \mathbf{s}^t) \geq \mathcal{C}(\mathbf{s})$ is satisfied if

$$(\mathbf{s} - \mathbf{s}^t)\left(\mathbf{K}(\mathbf{s}^t) - \mathbf{A}^T\mathbf{A} + \frac{\lambda}{4}\mathrm{diag}(\mathbf{s}^t)^{-\frac{3}{2}}\right)(\mathbf{s} - \mathbf{s}^t)^T \geq 0 \Rightarrow$$

$$(\mathbf{s} - \mathbf{s}^t)\left(\mathbf{K}'(\mathbf{s}^t) + \frac{\lambda}{2}\mathrm{diag}(\mathbf{s}^t)^{-\frac{1}{2}} + \frac{\lambda}{4}\mathrm{diag}(\mathbf{s}^t)^{-\frac{3}{2}}\right)(\mathbf{s} - \mathbf{s}^t)^T \geq 0 \tag{19}$$

where we have omitted $\mathcal{R}$ as it appears on both sides of the inequality and $\mathbf{K}'(\mathbf{s}^t)$ is defined as

$$\mathbf{K}'(\mathbf{s}^t) = \mathrm{diag}\left(\mathbf{A}^T\mathbf{As}^t./\mathbf{s}^t\right) - \mathbf{A}^T\mathbf{A} \tag{20}$$

Lee and Seung [12] proved the positive semidefiniteness of $\mathbf{K}'(\mathbf{s}^t)$. Due to the non-negativity of $\mathbf{s}$, the other two terms in Equation (19) are non-negative. This is due to the fact that the sum of two positive semidefinite matrices is also positive semidefinite. Further, Equation (19) holds by substituting Equation (16) into Equation (14). This results in the update rule

$$\begin{aligned}
\mathbf{s}^{(t+1)} &= \mathbf{s}^t - \nabla\mathcal{C}(\mathbf{s}^t)\mathbf{K}^{-1}(\mathbf{s}^t) \\
&= \mathbf{s}^t - \left(\mathbf{A}^T\mathbf{As}^t - \mathbf{A}^T\mathbf{x} + \frac{\lambda}{2}(\mathbf{s}^t)^{-\frac{1}{2}}\right).*\mathbf{s}^t./\left(\mathbf{A}^T\mathbf{As}^t + \frac{\lambda}{2}(\mathbf{s}^t)^{-\frac{1}{2}}\right) \\
&= \mathbf{s}^t.*\mathbf{A}^T\mathbf{x}./\left(\mathbf{A}^T\mathbf{As}^t + \frac{\lambda}{2}(\mathbf{s}^t)^{-\frac{1}{2}}\right)
\end{aligned}$$

Note that Equation (12) is the matrix form of the the update rule above. Thus, as long as the initial values of **A** and **S** are set in a strictly positive manner, the update rule above guarantees that the elements of the two matrices **A** and **S** remain non-negative. This makes the objective function in Equation (5) decrease monotonically at each iteration until convergence has been reached.

B. Implementation Issues

As mentioned earlier, the $L_{1/2}$-NMF is not a convex optimization problem with respect to both **A** and **S**. As a result, the rescaled gradient decent algorithm with the above update rules can only attain a local minimum. This implies that a number of factors will influence the final results. Firstly, the full additivity constraint of the end-member abundances can reduce the solution space of the optimization. Moreover, it should be noted that the full additivity constraint differs from the normalization of the columns of the matrix **A** often introduced in non-negative factorization approaches so as to avoid trivial solutions. Here, we employ a method akin to that in [33] where the data matrix **X** and the signature matrix **A** are augmented by a row of constants defined by

$$\mathbf{X}_f = \begin{bmatrix} \mathbf{X} \\ \delta \mathbf{1}_N^T \end{bmatrix} \quad \mathbf{A}_f = \begin{bmatrix} \mathbf{A} \\ \delta \mathbf{1}_K^T \end{bmatrix} \tag{21}$$

where δ controls the impact of the additivity constraint over the end-member abundances. The larger the δ, the closer the sum over the columns of **S** are to unity. In each iteration, these two matrices are taken as the input of the update rule of **S** given in Equation (12) as an alternative to **X** and **A**.

Note that the initialization of the signature matrix **A** can be computed by applying the end-member extraction method in [34] or using manually chosen data [35]. For the sake of easiness of implementation, **A** and **S** are both initialized by setting their entries to random values in the interval $[0, 1]$. Since the estimation of the number of end-members in the scene is crucial in the unmixing process, here we resort to the HySime algorithm [36]. We do this due to its reliability as an estimators for signal subspace dimensionality.

Here, we have adopted two stopping criteria for our iterative optimisation. The first of these is the maximum iteration number, which, in our experiments is set to be 3000. The second of these is the gradient difference of the cost function $\mathcal{C}$ between the current iteration and the starting value, i.e.

$$\|\nabla \mathcal{C}(\mathbf{A}^i, \mathbf{S}^i)\|_2^2 \leq \epsilon \|\nabla \mathcal{C}(\mathbf{A}^1, \mathbf{S}^1)\|_2^2$$

where ϵ is set to $\epsilon = 10^{-3}$ in our experiments. Once either of these criteria is met, the optimisation ends.

The value of the parameter λ is dependent on the sparsity of the material abundances. Since these abundances cannot be obtained a priori, we use a rough estimator for λ based on the sparseness criteria in [20]. This is given by

$$\lambda = \frac{1}{\sqrt{L}} \sum_l \frac{\sqrt{N} - \|\mathbf{x}_l\|_1 / \|\mathbf{x}_l\|_2}{\sqrt{N} - 1} \tag{22}$$

where $\mathbf{x}_l$ denotes the lth band in hyperspectral imagery.

In order to improve the robustness of the algorithm, not all elements in **S** are updated following the application of Equation (12). For those elements less than a predefined threshold, we omit the additional term corresponding to

the $L_{1/2}$-sparsity operator. In our experiments, the threshold is set as 10^{-4}. Note that the computational complexity at each iteration is linear with respect to the number of pixels in the scene, i.e. N, so the algorithm can be applied to hyperspectral imagery of medium and large sizes. For very large hyperspectral images, projected gradient algorithms, random block-wise methods, multi-layer processing and parallel processing schemes can be used [37], [38].

Finally, it should be noticed that our approach is quite general in nature. Indeed, other constraints can be added. For example, spatial information can be added in an akin manner to that applied in [15] to the L_1-NMF. This can provide an important advantage for hyperspectral unmixing.

The $L_{1/2}$-NMF based unmixing algorithm is summarized below.

Algorithm: $L_{1/2}$-NMF for Hyperspectral Unmixing

1) Estimate the number of K end-members using the HySime algorithm
2) Estimate the weight parameter λ according to the sparsity measure over $\mathbf{X}$.
3) Initialize $\mathbf{A}$ and $\mathbf{S}$ by randomly selecting entries in the interval [0, 1]. Rescale each column of $\mathbf{S}$ to unit norm.
4) Repeat

    a) Augment $\mathbf{X}$ and $\mathbf{A}$ to recover $\mathbf{X}_f$ and $\mathbf{A}_f$, respectively

    b) Do $\mathcal{C}_{old} = \mathcal{C}(\mathbf{A}_f, \mathbf{S})$

    c) Update $\mathbf{A}$ by applying Equation (11)

    d) Update $\mathbf{S}$ making use of Equation (12)

    e) Do $\mathcal{C}_{new} = \mathcal{C}(\mathbf{A}_f, \mathbf{S})$

  until the maximum number of iterations has been reached or $\|\mathcal{C}_{new} - \mathcal{C}_{old}\|_2^2 < \epsilon$

IV. EXPERIMENTS

Having presented our method in the previous section, we now turn our attention to demonstrating its utility for purposes of unmixing. Here, we employ synthetic and real-world data so as to evaluate the performance of the algorithms making use of the spectral angle distance (SAD) and the root mean squared error (RMSE). The SAD is used to compare the similarity of the kth end-member signature $\mathbf{A}_k$ and its estimate $\widehat{\mathbf{A}}_k$, which is defined as

$$\text{SAD}_k = \arccos\left(\frac{\mathbf{A}_k^T \widehat{\mathbf{A}}_k}{\|\mathbf{A}_k\| \|\widehat{\mathbf{A}}_k\|}\right) \tag{23}$$

On the other hand, the root mean square error (RMSE) is used to evaluate the the abundance estimates. The RMSE is defined as

$$\text{RMSE}_k = \left(\frac{1}{N} \mid \mathbf{S}_k - \widehat{\mathbf{S}}_k \mid^2\right)^{\frac{1}{2}} \tag{24}$$

where $\widehat{\mathbf{S}}_k$ is the ground-truth abundance matrix for the kth end-member.
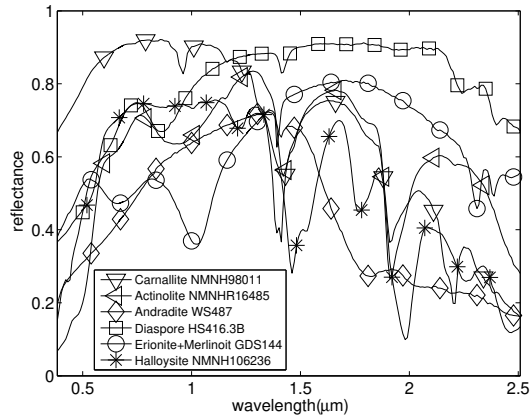
Fig. 1. Example spectral signatures from USGS used in our synthetic data experiments.

A. Synthetic Data

Firstly, we present a quantitative analysis of our method on synthetic data. Here, the proposed $L_{1/2}$-NMF algorithm is compared against three alternatives. These are the standard NMF in [12], the L_1-NMF [32] and the L_2-NMF method [13].

For our synthetic data experiments, ten spectral signatures are chosen from the United States Geological Survey (USGS) digital spectral library [39]. Figure 1 shows six example end-member signatures in them, and they will be used for all the following experiments in which the number of end-members is 6. The other four spectral signatures are Chlorite HS179.3B, Axinite HS342.3B Galena S26-39, Goethite WS220, and they will be used for the experiments that evaluate the unmixing methods to end-memeber number variation. To generate our synthetic data, we have computed ground-truth abundances in a similar manner to that in [35]. That is, we have departed from an image with size $z^2 \times z^2$ ($z \in \mathbb{Z}^+$) pixels, which is divided into $z \times z$ regions. Each region is initialized with the same type of ground cover, randomly selected as one of the end-members in the USGS dataset. We then use a $(z+1) \times (z+1)$ low pass filter on each pixel in the image to generate mixed pixels, and make the abundance variation smooth. Finally, we use a threshold θ ($0.6 \leq \theta \leq 1$) so as to make some pixels higher mixed. If a pixel whose abundance is larger than θ, this pixel is changed into a new mixed one made up of all end-members of equal abundances. This threshold parameter can be used to produce the synthetic data with different levels of sparseness, i.e., the smaller the θ, the higher mixed or called the less sparsity. The measure of average sparseness of all pixels in an imagery is defined as

$$sparse(\mathbf{A}) = \frac{1}{\sqrt{K}} \sum_l \frac{\sqrt{K} - \|\mathbf{A}_k\|_1/\|\mathbf{A}_k\|_2}{\sqrt{K} - 1} \tag{25}$$

where $\mathbf{A}_k$ denotes the abundances on k^{th} end-menber.

For our sensitivity study, we have used zero-mean white Gaussian noise, which, when substituted into Equation
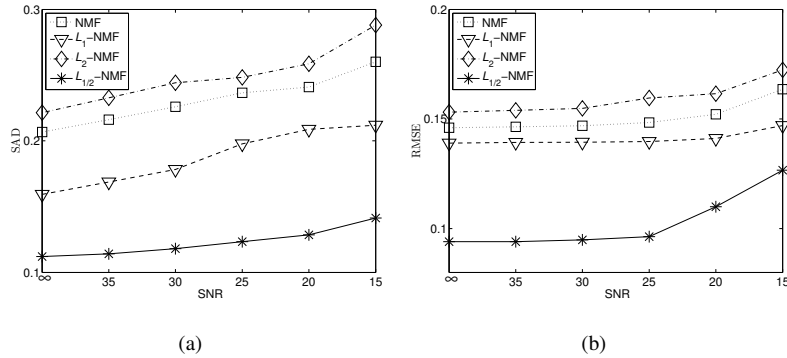
Fig. 2. Results on synthetic data: SAD (a) and RMSE (b) as a function of SNR.

(1), yields the following signal-to-noise ratio (SNR)

$$\text{SNR} = 10\log_{10}\frac{E[(\mathbf{As})^T(\mathbf{As})]}{E[\mathbf{e}^T\mathbf{e}]} \tag{26}$$

where $E[\cdot]$ denotes the expectation operator.

Note that, from the synthetic data generation process above, its natural to perform a quantitative analysis of the four methods, i.e. our proposed $L_{1/2}$-NMF unmixing algorithm and the other three alternatives, with respect to the SNR, the sparseness, the image size and the number of end-members. To this end, we commence by showing, in Figure 2 the RMSE and SAD levels as a function of the SNR for the interval $(\infty,\ldots,15)$ in steps of 5 dBs for six end-members, i.e. $K = 6$, $\theta = 0.7$ and $z = 7$. As expected, the decrease in the SNR has a detrimental effect in the performance for all the four algorithms. From the figure, we can see that the L_2-NMF delivers the worst results for both the SAD and the RMSE. The performance of the L_1-NMF is slightly better than that of NMF. Meanwhile, our $L_{1/2}$-NMF not only provides the best SAD, but its also robust to noise corruption by yielding the smallest RMSE.

The second experiment is to evaluate the performance of $L_q(0 < q \leq 1)$ regularizer to sparseness, so L_2-NMF is removed, and $L_{1/4}$ and $L_{3/4}$-NMFs are added for comparison. Figure 3 shows the plot of the unmixing results with different sparseness levels of the end-member abundances. Here the parameters are set as $K = 6$, $z = 7$, and θ controlling the sparseness levels. From figure 3(a), we can see that the SAD of L_1-NMF is basically not affected by the variation of the sparsity, while those of the other three L_q-NMF $(0 < q < 1)$ decrease with the increase of sparseness, supporting the assumption that L_1 regularizer cannot imposing any sparsity on the solution when full additivity constraint is enforced, which is the reasons why L_q-NMF $(0 < q < 1)$ should be introduced to spectral unmixing. In figure 3(b), along with the increase of the sparsity, the performance of L_1-NMF decreases, and the L_q-NMF $(0 < q < 1)$ methods show the similar variation with SAD. Figure 3(c) shows the true sparseness vs the estimated sparseness. Clearly, the sparsity obtained by L_1-NMF is much lower than the true sparseness along with the increase of the abundance sparsity. On the contrary, the estimated sparsities by L_p-NMF $(0 < p < 1)$ are close to the true ones. Furthermore, from figures **??**, we found that the results of $L_{1/2}$-NMF and $L_{1/2}$-NMF are very similar, supporting another important assumption that the sparsity of the $L_q(1/2 \leq q < 1)$ solution increases as q decreases, whereas the sparsity of the solution for $L_q(0 < q \leq 1/2)$ is almost indifferent with respect to q, which
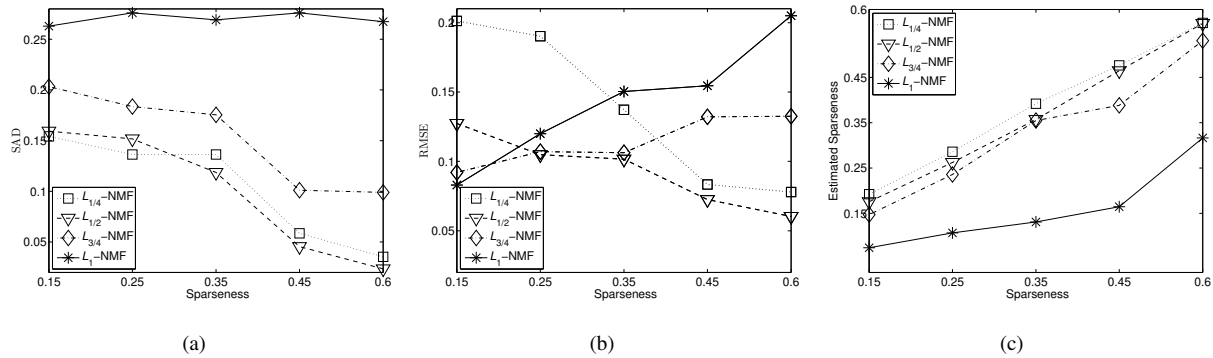
Fig. 3. Results on synthetic data: SAD (a) and RMSE (b) as function of different sparsity of the abundances. (c) the true sparsity vs the estimated sparsity obtained by the four methods.
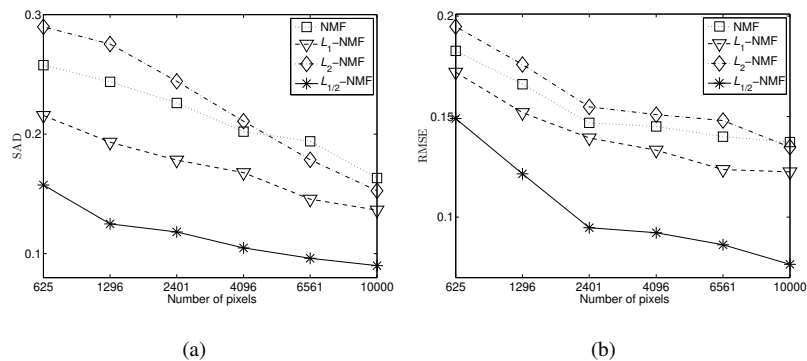

Fig. 4. Results on synthetic data: SAD (a) and RMSE (b) as a function of the number of pixels in the scene.

is a reason why $q = 1/2$ is a better choice among $0 < q < 1$ for spectral unmixing.

We now examine the effect of varying the number of pixels on our method and the alternatives. We set the number of pixels to $625, 1296, \dots, 10000$, which corresponds to $z = 5, 6, \dots, 10$. Again, we have used six end-members and set the SNR = 30db and $\theta = 0.7$. From Figure 4, we can see that the performances of the four methods become to better as the size of the scene increases. This is expected, since a large size of training examples makes the solution space more stable and constrained, increasing the likelihood of finding an optimal or near-optimal solution.

Finally, we examine the performance of the four methods to end-member number variation and their accuracy with respect to the estimated number of end-members. To do this, we vary the number of end-members from $K = 3$ to $K = 10$. Figure 5 shows that the performance of the four methods when the SNR is 30db, $\theta = 0.7$ and $z = 7$. Note that the performance decays as the number of end-members present in the scene increases. Indeed, when $K = 3$, the four methods are comparable. As the number of end-members increases, the difference in performance between the algorithms becomes more apparent, with the $L_{1/2}$-NMF consistently producing the best results. Note that, in practice, overestimation of the number of end-members present in the scene is often a crucial problem in hyperspectral unmixing. Hence, we also study the robustness of the algorithm to the overestimation of the number of end-members. Figure 6 shows the plot of the unmixing results when the number of end-members
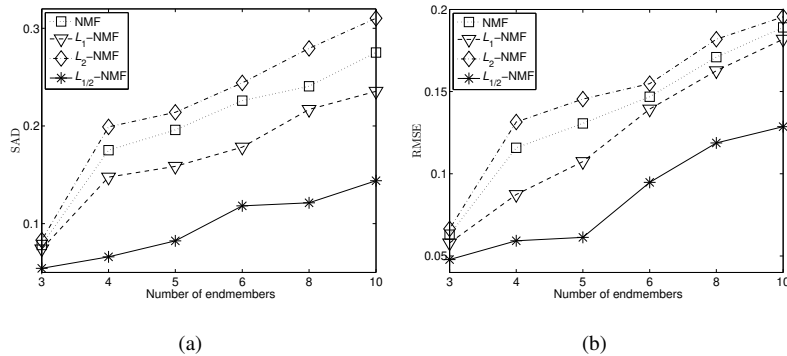
Fig. 5. Results on synthetic data: SAD (a) and RMSE (b) as function of the number of end-members.
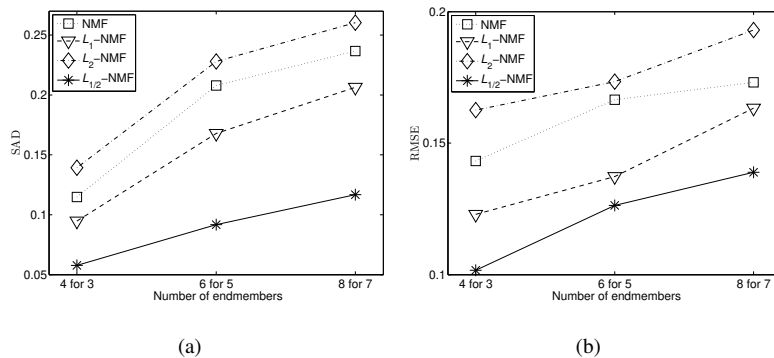

Fig. 6. Results on synthetic data: SAD (a) and RMSE (b) as function of the number of end-members overestimated by 1.

is overestimated by 1. As before, the results obtained by our $L_{1/2}$-NMF method are more robust than those of the other three alternatives. This is in-line with the our other experiments on synthetic data, where it can be observed that $L_{1/2}$-NMF is a better alternative than the NMF, the L_1-NMF and the L_2-NMF.

At last, we discuss the computational complexity of the L_p-NMF ($0 < p < 1$) methods. The computational complexity of the $L_{1/2}$-NMF is similar to that of standard NMF, except the computation load of $S^{-1/2}$, which is known as $O((KN)^2)$. Hence, the computational complexity of $L_{1/2}$-NMF is $O(LKN + (KN)^2)$ of each iteration. Furthermore, the computational costs of other $L_q (0 < q < 1)$ NMF are usually larger than $L_{1/2}$-NMF, which is also a reason that we select $q = 1/2$.

B. Experiments on Real-word Data

Now we present the results of applying our $L_{1/2}$-NMF method to real-world data. Here, we have used two data sets which cover both, an urban scene and an image that provides a regional geologic context.

The first real-world data set to be used is obtained from the Urban HYDICE hyperspectral image. The image depicts the scene in Figure 7 and is of size 307×307. The image is composed of 210 spectral channels with spectral resolution of 10nm acquired in the 400nm and 2500nm region. After low SNR bands are removed (channels $1-4$, 76, 87, $101-111$, $136-153$, and $198-210$), only 162 bands remain (i.e., $L = 162$). There are four distinct targets

Fig. 7.   Urban HYDICE hyperspectral dataset at band 80.

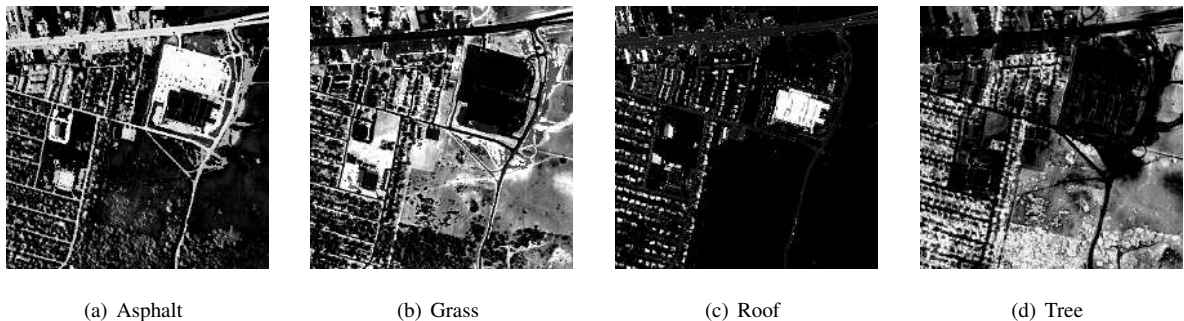(a) Asphalt   (b) Grass   (c) Roof   (d) Tree

Fig. 8.   Ground-truth abundance maps for four targets in the Urban HYDICE hyperspectral data.

of interest: asphalt, grass, roof and tree. Figure 8 displays the ground truth for the abundance fractions of the end-members. In these images, and from now on, pure black denotes that the fraction of a certain object in the given pixel is null, while pure white denotes 1.

To evaluate the effectiveness of the proposed algorithm, we have compared our method against the L_1-NMF. We have done this since the latter performs the best among the three alternative methods examined in our previous experiments on synthetic data. Meanwhile, three state-of-the-art methods related to the proposed algorithm are compared, i.e., VCA [3], MVC-NMF [35], PSNMFSC [15] and SISAL [40]. VCA is a popular geometrical based approach with the presence of pure pixels. MVC-NMF and PSNMFSC are both constrained NMF methods, and the first uses minimum volume as constraint while the later uses piecewise smoothness and sparseness. SISAL is the abbreviation of Simplex Identification via variable Splitting and Argumented Lagrangian, which can deal with outliers and is very efficient from the computational aspect.

Figures 9 illustrate the separated abundance fractions for each end-member as delivered by our $L_{1/2}$-NMF. Meanwhile, Figure 10 display the estimated $L_{1/2}$-NMF end-member signatures with respect to the USGS library spectra. It can be found that the results of $L_{1/2}$-NMF is very close to the real end-member signatures and abundances.
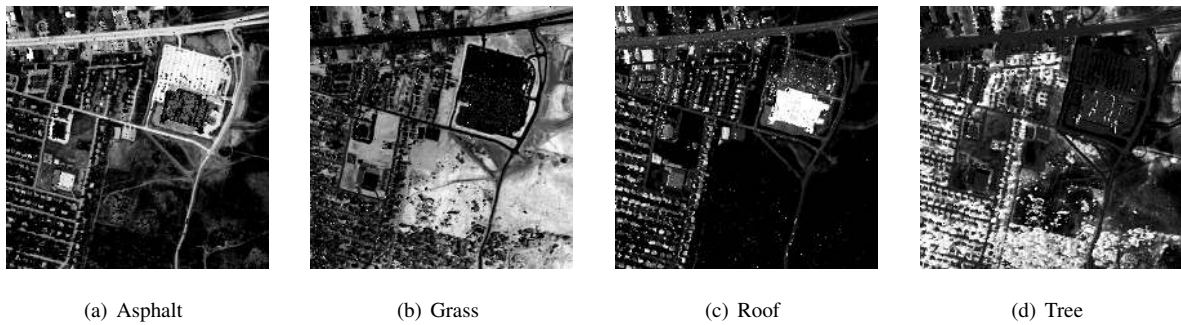
(a) Asphalt   (b) Grass   (c) Roof   (d) Tree

Fig. 9. Urban HYDICE results: Abundance maps estimated using $L_{1/2}$-NMF for the four targets.

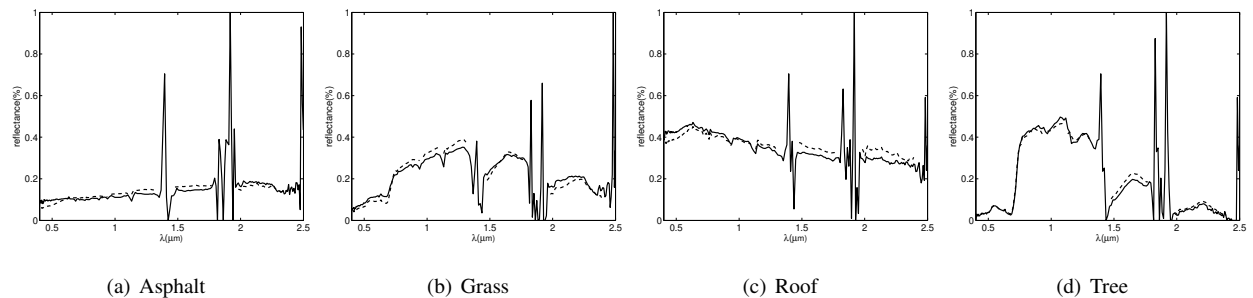(a) Asphalt   (b) Grass   (c) Roof   (d) Tree

Fig. 10. Urban HYDICE results: Comparison of the USGS library spectra (solid line) with the signatures extracted by $L_{1/2}$-NMF (dotted line).

Table I gives the SAD criterioa of all six algorithms and their corresponding standard variances (each method repeatedly runs 10 times). From this table, it can be seen that the results obtained by our $L_{1/2}$-NMF are better than the other algorithms.

In [41], the road is further divided into asphalt road and concrete, and roof is divided into roof #1 and roof #2/shadow and concrete road. In most cases, the exact number of end-members cannot be obtained or even it does not exist because the assumed number of end-members is dependent on the analysis scale. In order to test our algorithms, we also do the same experiments with six end-members. From the Figures 11, 12, 13 and Table II, we also can find that the results obtained by our $L_{1/2}$-NMF are better than the other algorithms.

Now we turn our attention to the second real-world dataset. This is an image acquired by the AVIRIS sensor over Cuprite in Souther Nevada. It is a regional scene which contains an abundant supply of minerals [42]. In recent years, the Cuprite data set has been widely used for hyperspectral unmixing research [3], [35]. Figure 14 displays the 80th band as a subimage of the original data with size 250×190. For our experiments we have removed low SNR and water-vapor absorption bands (1-2, 104-113, 148-167 and 221-224), which yields 188 bands out of the original 224.

According to [3], there are 14 types of minerals present in the scene. Its worth mentioning that variants of the same mineral with slightly different spectra are not considered as dissimilar end-members and, hence, the number of

TABLE I

SAD AND THE STANDARD VARIANCE (%) RESULTS ON THE URBAN HYDICE DATA WITH 4 ENDMEMBERS.

	$L_{1/2}$-NMF	L_1-NMF	MVCNMF	VCA	SISAL
Asphalt	**0.1352± 0.88**	0.3347± 0.86	0.5073± 5.65	0.3091± 1.10	0.1868± 3.71
Grass	**0.2097 ± 2.65**	0.2282±4.82	0.2908±2.25	0.3196±8.40	0.2598±1.06
Tree	**0.0658±3.05**	0.1436±0.56	0.2124±5.97	0.2110±2.57	0.1724±2.77
Roof	0.2441±6.03	0.4664±0.35	0.2783±5.37	0.7619±0.20	**0.2097±0.52**
Mean	**0.1637±3.15**	0.2932±1.65	0.3222±4.81	0.4004±3.07	0.2072±2.02

(a) Asphalt Road

(b) Grass

(c) Tree
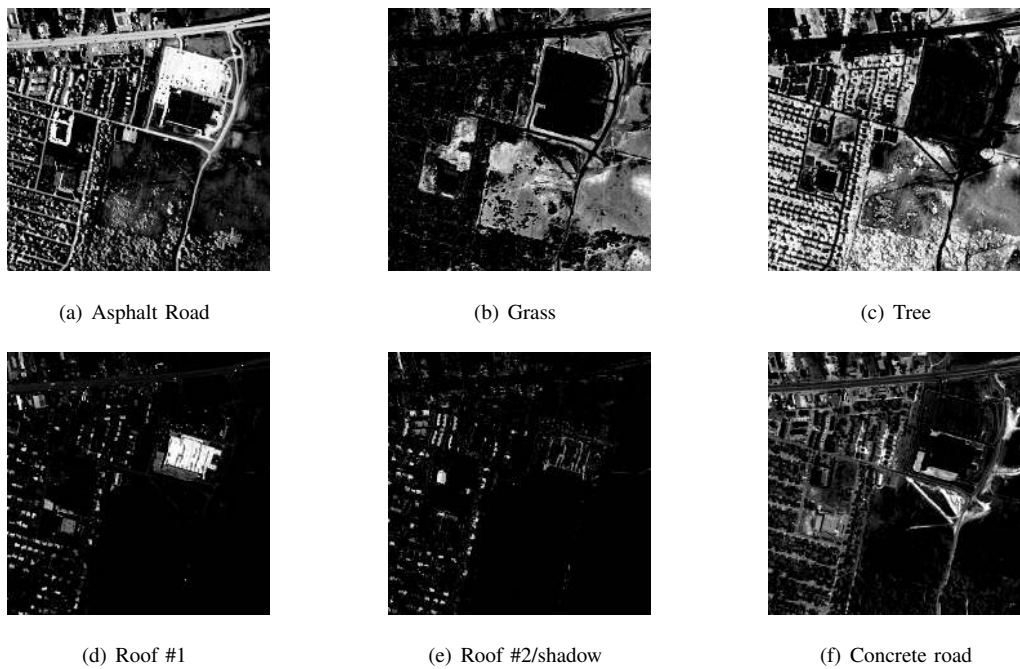
(d) Roof #1

(e) Roof #2/shadow

(f) Concrete road

Fig. 11.   Ground-truth abundance maps for six targets in the Urban HYDICE hyperspectral data.

end-member is set to 10, i.e., $K = 10$. In Figure 15, we compare the estimated $L_{1/2}$-NMF end-member signatures with the USGS library spectra. Clearly, the extracted signatures are in good accordance with the USGS library spectra. Table III quantifies the similarity of the recovered spectra using the SAD criterion. For most of materials in the image, the SAD of our method is lowest, which shows our algorithms is better than the others.

V. CONCLUSIONS

In this paper, we have extended NMF-based hyperspectral unmixing methods by incorporating the $L_{1/2}$ sparseness constraint over the end-member abundances. In contrast with previous approaches, which used the L_p, $(p \geq 1)$ regularizer, our $L_{1/2}$-NMF produces sparser unmixing results with the end-member spectra and abundance maps being more accurately recovered. We have also presented an effective multiplicative iterative algorithm, which
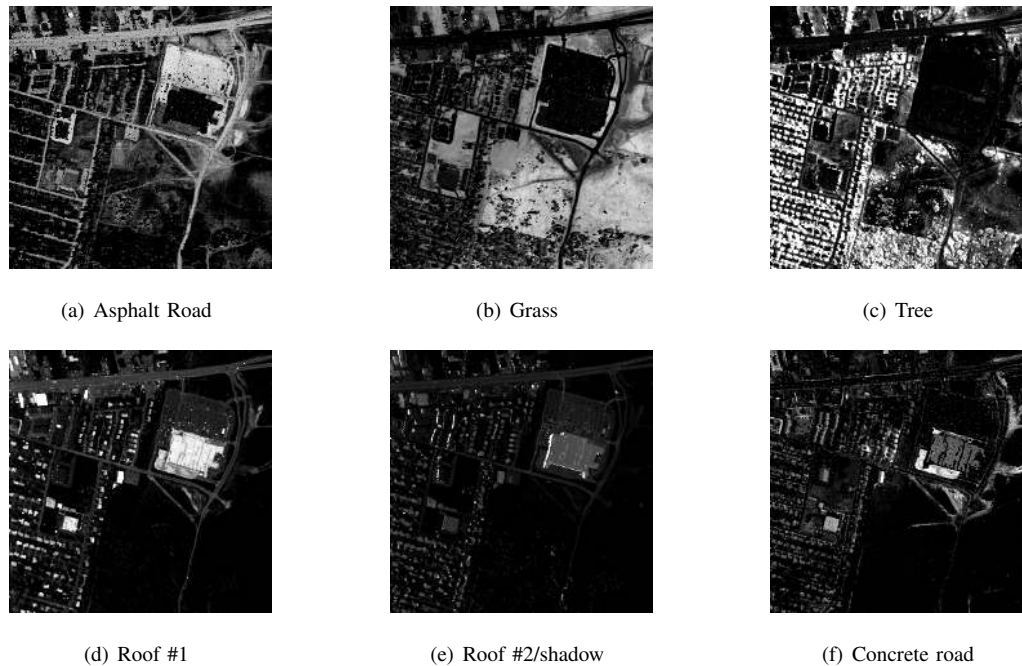
(a) Asphalt Road   (b) Grass   (c) Tree

(d) Roof #1   (e) Roof #2/shadow   (f) Concrete road

Fig. 12. Urban HYDICE results: Abundance maps estimated using $L_{1/2}$-NMF for the six targets.

TABLE II

SAD AND THE STANDARD VARIANCE (%) RESULTS ON THE URBAN HYDICE DATA WITH 6 ENDMEMBERS.

	$L_{1/2}$-NMF	L_1-NMF	MVCNMF	VCA	SISAL
Asphalt road	0.2302± 1.03	0.3739± 1.81	0.4978± 5.49	0.3304± 7.70	**0.2263± 2.03**
Grass	**0.2692 ± 2.95**	0.3772±3.40	0.3334±4.57	0.5104±3.06	0.3655±3.29
Tree	**0.0414±1.14**	0.1523±1.45	0.1727±4.99	0.3108±3.79	0.2513±1.59
Roof #1	**0.1000±3.56**	0.7134±2.66	0.2574±4.35	0.9529±6.56	0.4339±5.53
Roof #2/shadow	**0.2617±4.27**	0.5849±0.97	0.4389±5.06	0.5409±3.22	0.3373±4.65
Concrete road	**0.1570±0.83**	0.5136±0.60	0.2882±4.34	0.3721±0.46	0.3119±8.84
Mean	**0.1766±2.30**	0.4526±1.82	0.3314±4.80	0.5029±4.13	0.3210±4.32

estimates the end-member signatures and abundances by making use of a rescaled gradient descend. We have illustrated the advantages of our unmixing method on synthetic and real-world data and compared our method to a number of alternatives, i.e., NMF, L_1-NMF and L_2-NMF, VCA, MVC-NMF, PSNMFSC and SISAL. The experimental results consistently show that $L_{1/2}$-NMF exhibits better performance. This is particularly true in the presence of noise corruption and low end-member purity levels.

We would like to note that the $L_{1/2}$-NMF method presented here is quite general in nature and can be readily applied to other setting in which nonnegative sparse matrix factorization is a valuable computational tool. Furthermore, the method presented here can easily incorporate constraints presented elsewhere in the literature. In
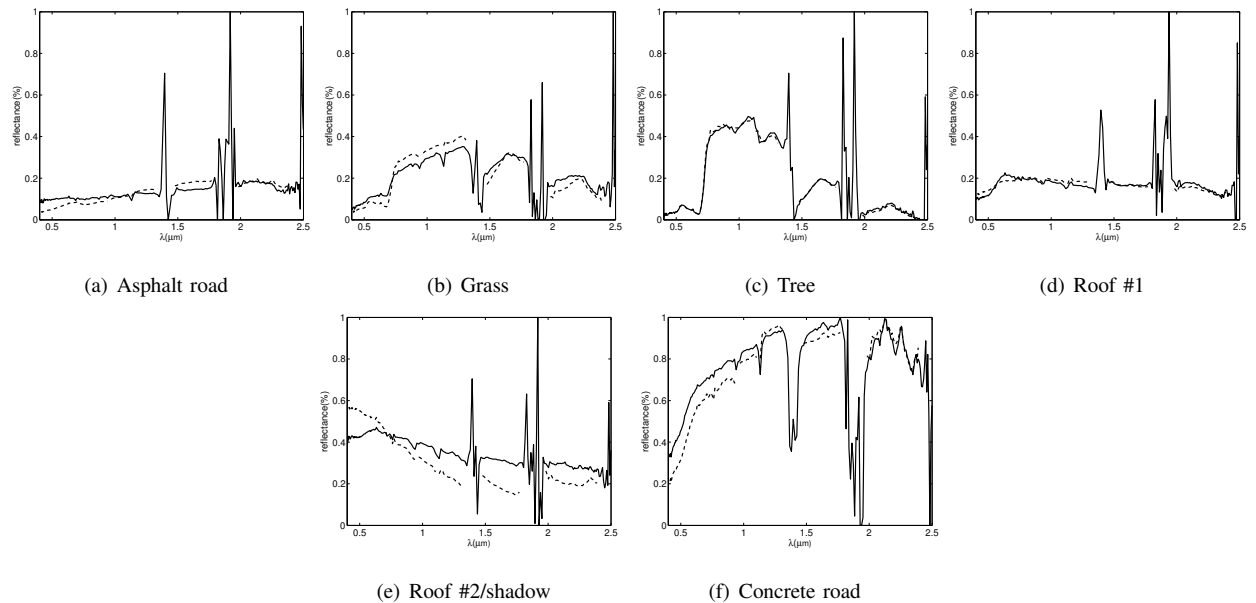
(a) Asphalt road   (b) Grass   (c) Tree   (d) Roof #1

(e) Roof #2/shadow   (f) Concrete road

Fig. 13. Urban HYDICE results: Comparison of the USGS library spectra (solid line) with the signatures extracted by $L_{1/2}$-NMF (dotted line) with 6 endmembers.
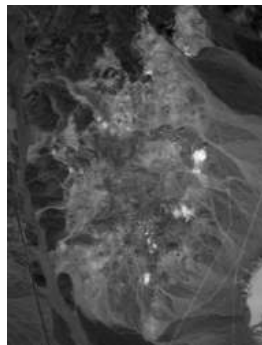

Fig. 14. The AVIRIS Cuprite image (band 80).

future research, we aim at studying extensions of $L_{1/2}$-NMF for unmixing such as wavelet-subspace, Markovian formulations based upon spatial consistency constraints and $L_{1/2}$ with non-negative tensor factorization. Another possibility is to develop more effective estimation algorithms for $L_{1/2}$-NMF and their extensions to robust statistics so as to achieve greater levels of robustness to noise. In this regard, the use of Bayesian estimation methods is worth considering.

REFERENCES

[1] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 55–78, 2003.

[2] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry V*, 1999, pp. 266–275.
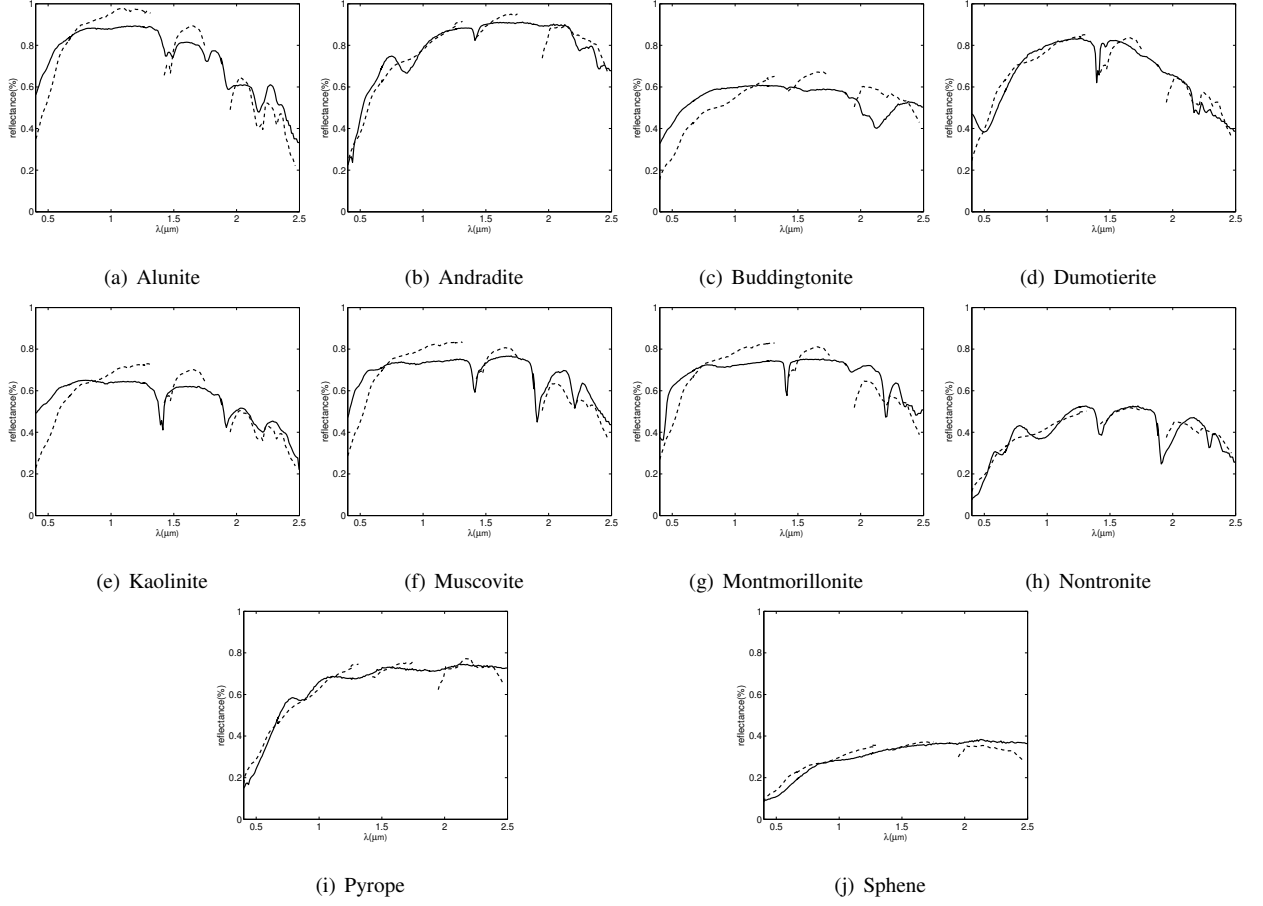
(a) Alunite   (b) Andradite   (c) Buddingtonite   (d) Dumotierite

(e) Kaolinite   (f) Muscovite   (g) Montmorillonite   (h) Nontronite

(i) Pyrope   (j) Sphene

Fig. 15.   Results on the AVIRIS Cuprite image: Comparison of the USGS library spectra (solid line) with the signatures extracted by our $L_{1/2}$-NMF (dotted line).

TABLE III

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

	$L_{1/2}$-NMF	L_1-NMF	MVCNMF	VCA	SISAL
Alunite	0.1660± 0.61	0.2874± 6.83	0.1421± 5.92	**0.1016± 3.19**	0.2103± 5.63
Andradite	**0.0545 ± 0.14**	0.2380±0.87	0.1488±5.11	0.0715±3.18	0.0869±4.57
Buddingtonite	0.1626±0.25	0.3448±8.86	0.2058±4.44	**0.1317±1.99**	0.1413±4.98
Dumotierite	**0.0918±1.27**	0.2112±4.87	0.1553±3.90	0.0988±1.30	0.1221±4.77
Kaolinite	**0.1441±4.46**	0.3793±3.46	0.1979±4.15	0.2418±4.85	0.2087±3.97
Montmorillonite	**0.1206±2.26**	0.3014±4.75	0.1339±9.87	0.1357±4.52	0.1605±4.03
Muscovite	0.1292±2.39	0.3801±5.01	**0.0914±9.28**	0.1349±4.01	0.1398±4.58
Nontronite	**0.0850±0.08**	0.2292±0.92	0.1779±8.03	0.0920±1.81	0.0887±3.89
Pyrope	**0.0596±0.39**	0.3188±0.43	0.2078±7.71	0.1257±3.77	0.1482±3.96
Sphene	0.1002±0.28	0.2718±5.23	0.1439±3.73	0.0794±5.16	**0.0713±2.26**
Mean	**0.1114±1.21**	0.2962±4.12	0.1585±6.21	0.1213±3.38	0.1378±4.26

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

[4] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.

[5] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 4418–4432, 2009.

[6] J. Chen, X. Jia, W. Yang, and B. Matsushita, "Generalization of subpixel analysis for hyperspectral data with flexibility in spectral similarity measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2165–2171, 2009.

[7] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3867–3879, 2007.

[8] Y. Qian and Q. Wang, "Noise-robust subband decomposition blind signal separation for hyperspectral unmixing," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, 2010.

[9] J. Ye and Q. Li, "Lda/qr: An efficient and effective dimension reduction algorithm and its theoretical foundation," *Pattern recognition*, vol. 37, pp. 851–854, 2004.

[10] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*.   Kluwer Academic/Plenum Publishers, New York, 2003.

[11] P. Paatero and U. Tapper, "Positive matrix factorization:a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111–126, 1994.

[12] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[13] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 29–47, Jul. 2006.

[14] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: applications to exploratory multi-way data analysis and blind source separation*.   Wiley, 2009.

[15] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 161–173, 2009.

[16] M. D. Iordache, A. Plaza, and J. Bioucas-Dias, "Recent developments in sparse hyperspectral unmixing," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2010.

[17] D. Thompson, R. Castao, and M. Gilmore, "Sparse superpixel unmixing for exploratory analysis of CRISM hyperspectral images," in *Proceedings of the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2009, pp. 1–4.

[18] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[19] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[20] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, November 2004.

[21] A. P.-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. P.-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, March 2006.

[22] A. Zare and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 256–260, 2008.

[23] B. Liu, S. Chen, M. Qian, and C. Zhang, "Sparse norm-regularized reconstructive coefficients learning," in *Proceedings of the Ninth IEEE International Conference on Data Mining*, 2009, pp. 854–859.

[24] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Algo-rithms and Technologies for Multispectral, Hyperspectral, andUltraspectral Imagery*, S. S. Shen and P. E. Lewis, Eds., vol. 7334, 2009, pp. 73 341M–3341M–9.

[25] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, pp. 1348–1360, 2001.

[26] J. Fan and P. Heng, "Nonconcave penalty likelihood with a diverging number of parameters," *The Annals of Statistics*, vol. 32, pp. 928–961, 2004.

[27] Z. Xu, H. Zhang, Y. Wang, and Y. L. X.Y. Chang, "$L_{1/2}$ regularizer," *Science in China, series F*, vol. 53, pp. 1159–1169, 2010.

[28] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*. MIT Press, 2001, pp. 556–562.

[29] G. A. Shaw and H. Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.

[30] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[31] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[32] P. O. Hoyer, "Non-negative sparse coding," in *Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, Martigny, Switzerland, 2002, pp. 557–565.

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

[34] J. M. P. Nascimento and J. M. B. Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. IGARSS 2007*, 2007, pp. 4033–4036.

[35] L. D. Miao and H. R. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[36] J. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[37] A. Cichocki, R. Zdunek, and S.-i. Amari, "Nonnegative matrix and tensor factorization," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 142–145, 2008.

[38] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3373–3376.

[39] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," U.S. Geol. Surv. Open File Rep. 93-592, 1993.

[40] J. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. First IEEE WHISPERS*, Aug. 2009, pp. 1–4.

[41] X. Liu, W. Xie, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, will be published.

[42] G. A. Swayze, R. L. Clark, S. Sutley, and A. J. Gallagher, "Ground-truthing AVIRIS mineral mapping at cuprite, nevada," in *Summaries of the 3rd Annual JPL Airborne Geosciences Workshop*, vol. 1, 1992, pp. 47–49.