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. Unfortunately, the L_1 regularizer cannot enforce further sparsity when the full additivity constraint of material abundances is used, 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 a better choice among $L_q(0 < q < 1)$ regularizers. We propose an iterative estimation algorithm for $L_{1/2}$-NMF, which provides sparser and more accurate results than those delivered using the L_1 norm. We illustrate the utility of our method on synthetic and real hyperspectral data and compare our results to those yielded by other state-of-the-art methods.

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,

Y. Qian is with the Institute of Artificial Intelligence, College of Computer Science, Zhejiang University, Hangzhou, P.R. China. He is supported by Australia China Special Fund for Science and Technology Cooperation No.61011120054

S. Jia is with the Texas Instruments DSPs Laboratory and College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, P.R. China. He is supported by National Natural Science Foundation of China No.60902070. Corresponding author.

J. Zhou and A. Robles-Kelly are with Canberra Research Laboratory, NICTA, PO BOX 8001, Canberra, ACT 2601, Australia. They are also with College of Engineering and Computer Science, The Australian National University, Canberra, ACT 0200, Australia. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

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]. Many hyperspectral unmixing methods have been proposed in recent years. These include N-FINDR [2], vertex component analysis (VCA) [3], independent component analysis (ICA) [4], alternating projected subgradients (APS) [5], minimum volume based algorithms [6], [7], [8], [9], [10] and flexible similarity measures [11].

Most of these methods assume a linear spectral mixture model for the unmixing process. If the number and signatures of end-members 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 [12]. Since the source signals are normally independent from one another in some specific frequencies, it can be assumed that the subcomponents of the sources are mutual independent, which leads to subband ICA [13].

From the linear algebra point of view, BSS is a constrained matrix factorization problem that has found numerous applications in feature and signal extraction [14]. 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 [15]. 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.

Non-negative matrix factorization (NMF) [16], [17], which decomposes the data into two nonnegative matrices, is a natural solution to the nonnegativity constraint [18]. 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 [19]. However, the solution space of NMF is very large if no further constrains are considered. This, added to the fact that the cost function is not convex, makes 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 [19]. Researchers have also tried improving NMF based unmixing methods by imposing further constraints [18], [20], [21]. In [22], Donoho and Stodden analyzed the assumptions required for NMF to generate unique solutions and lead to a well-defined answer. More recently, sparsity constraints have gained much attention since they allow exploiting the notion that most of the pixels are mixtures of only a few of the end-members in the scene [23], [24]. 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 end-members.

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 solution of the L_0 regularizer is an NP hard optimization problem that cannot be solved in practice. The L_2 regularizer, on the other hand, generates smooth but not sparse results [25]. In general, the L_1 regularizer is the most popular choice for achieving sparsity of the abundance matrix [26], [27], [28].

More recent works on this topic include the semi-supervised algorithms based on sparse regression in [29], [24]. These methods assume that the pixel signature can be expressed in the form of linear combinations of a number of pure spectral signatures known in advance. As the library of pure spectral signatures contain more possible sources than those actually present in the scene. The L_1 regularizer is used to favor sparse solutions where only a subset of signatures is selected as end-members. Guo *et al* [30] also used a sparse regression model to estimate the abundances, while the end-members are extracted using the N-FINDR algorithm . In [31], sparsity-promoting priors are used with an extension of the iterated constrained end-member (ICE) algorithm [32] to determine the number of end-members, in which the sparsity is achieved by a zero-mean Laplacian distribution akin to the L_1 regularizer.

Despite the importance of sparsity in hyperspectral unmixing, and the wide use of L_1 regularizer, the sparsity property of the regularizers and its influence on the unmixing performance has not been thoroughly investigated. Recently, the properties of fractional $L_q(0 < q < 1)$ regularizers have been studied in [33] and [34]. The $L_q(0 < q < 1)$ regularizers give sparser solutions than their L_1 counterpart. Furthermore, the sparsity of the $L_q(1/2 \leq q < 1)$ solution increases as q decreases, whereas the sparsity of the solution for $L_q(0 < q \leq 1/2)$ does not overly change with respect to q. Xu *et al* [35] have shown that the $L_{1/2}$ regularizer is an unbiased estimator which imposes strong sparsity upon the minimisation problem at hand.

In this paper, we introduce the $L_{1/2}$ regularization into NMF, which we name $L_{1/2}$-NMF, so as to enforce the sparsity of end-member abundances. The $L_{1/2}$-NMF presented here is implemented through the multiplicative update algorithm in [36], which is an iterative application of a rescaled gradient descent approach so as to ensure convergence. In our approach, the full additivity constraint is naturally embedded in the parameter update process. The experiments on synthetic and real-world hyperspectral data demonstrate the effectiveness of our $L_{1/2}$-NMF approach. We also generalize the method to L_q-NMF for $0 < q < 1$, and discuss the sparsity imposed by the regularizers upon the unmixing task.

Our contribution is, hence, to introduce a novel NMF method which recovers a sparse solution to the unmixing problem using an optimization algorithm that guarantees stable convergence to a local minimum. We show that the L_1 regularizer is not a good choice in enforcing the sparsity of hyperspectral unmixing and propose the use of $L_q(0 < q < 1)$ NMF instead. We analyze the behavior of these regularizers and point out that $q = 1/2$ is a sound option to the problem. The paper also provides a link between the sparsity-constrained NMF and minimum volume methods.

The rest of the paper is organized as follows. In Section II, we introduce the linear spectral mixture model and brief background to the nonnegative matrix factorization. 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. Section IV extends the algorithm to L_q-NMF for $0 < q < 1$, and discusses some important factors related to the proposed method. Results on synthetic and real-world data are reported in Sections V-A and Section V-B, respectively. Finally, Section VI draws conclusions and gives suggestions on future research.

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 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

Classical linear mixing is a way to represent the spectrum of a pixel of L wavelength-indexed bands in the observed scene based upon K end-member abundances. It 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 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 minimizing the difference between $\mathbf{X}$ and $\mathbf{AS}$ and enforcing non-negativity on $\mathbf{A}$ and $\mathbf{S}$. Such difference is often measured using the Euclidean distance, relative entropy, or Kullback Leibler divergence. The loss function for NMF based upon the Euclidean distance is as follows

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

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 [19]. 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 representations similar to those found in primary visual processing [37]. The importance of sparsity is also illustrated in compressive sensing [38]. Likewise, sparsity is an intrinsic property of hyperspectral data. In most cases, the abundance distribution of any end-member does not apply to the whole scene. This implies that the mixed pixel is usually the superposition of only a few end-members, but not all those present in the scene. That is, for each end-member, its abundance is localized with a degree of sparseness.

For this reason, 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 regularization 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 for training samples 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 does not enforce the 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, which is an alternative to the L_1 counterpart. As mentioned earlier, the $L_{1/2}$ regularizer is a sparsity-promoting function [35]. 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 k^{th} end-member at the n^{th} pixel in the image.

III. Algorithm for $L_{1/2}$-NMF based Unmixing

We now focus on achieving a factorization solution compliant with the additivity constraint over the end-members. 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 enclosing 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 others remain 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, which belong to the majorization-minimization family, have relatively low complexity and can overcome some of these drawbacks [39]. Lee and Seung derived a multiplicate update rule for standard NMF whose convergence has been proved [36]. When applied to Equation (3) this multiplicative update becomes

$$\mathbf{A} \leftarrow \mathbf{A}.*\mathbf{XS}^T./\mathbf{ASS}^T \tag{7}$$

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

An extension of the above multiplicative rule for L_1-NMF was later developed by Hoyer [40] as follows

$$\mathbf{A} \leftarrow \mathbf{A}.*\mathbf{XS}^T./\mathbf{ASS}^T \tag{9}$$

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

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 [40] can be achieved by a modification of the above multiplicative update rules as follows

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

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

Likewise, where $\mathbf{S}^{-\frac{1}{2}}$ is given by the sum over the element-wise square root for each entry in the matrix $\mathbf{S}$. If any entry in $\mathbf{S}$ is zero, a very small value will be added so as to avoid trivial solution.

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 only address 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

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

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 [40], 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

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

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

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

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

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

Lee and Seung [17] 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

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

Note that Equation (12) is the matrix form of the the update rule above. Thus, as long as the initial values of $\mathbf{A}$ and $\mathbf{S}$ are set in a strictly positive manner, the update rule guarantees that the elements of the two matrices $\mathbf{A}$ and $\mathbf{S}$ 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 $\mathbf{A}$ and $\mathbf{S}$. 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 result. 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 $\mathbf{A}$ often introduced in non-negative factorization approaches so as to avoid trivial solutions. Here, we employ a method akin to that in [41] where the data matrix $\mathbf{X}$ and the signature matrix $\mathbf{A}$ 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 $\mathbf{S}$ are to unity. In each iteration, these two matrices are taken as the input of the update rule of $\mathbf{S}$ given in Equation (12) as an alternative to $\mathbf{X}$ and $\mathbf{A}$.

Note that the initialization of the signature matrix $\mathbf{A}$ can be computed by applying the end-member extraction method in [42], or using manually chosen data [43]. For the sake of easiness of implementation, $\mathbf{A}$ and $\mathbf{S}$ 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 [44]. We do this due to its reliability as an estimators for signal subspace dimensionality.

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 3000. The second one 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 an estimator for λ based on the sparseness criteria in [26]. 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 l^{th} band in hyperspectral imagery.

In order to improve the robustness of the algorithm, not all elements in $\mathbf{S}$ 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 to 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 [19].

Finally, it should be noted that our approach is quite general in nature. Indeed, other constraints can be added. For example, spatial information can be added to the $L_{1/2}$-NMF in a manner akin to that applied in [20]. Our $L_{1/2}$-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. Method Extension and Discussions

A. NMF with L_q regularizer

To generalize the proposed algorithm to L_q-NMF for $0 < q < 1$, we note that, in theory, all L_qs for $0 < q < 1$ can be used for sparsity NMF. Extending the proposed method to NMF with L_q regularizer is straightforward. It follows the definition and algorithm in sections II-C and III-A. More specifically, the L_q-NMF model for unmixing is given by

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_2^2 + \lambda\|\mathbf{S}\|_q \tag{23}$$

where

$$\|\mathbf{S}\|_q = \sum_{k,n=1}^{K,N} \mathbf{s}_n(k)^q \tag{24}$$

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

The multiplicative update rules to solve $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{S})$ are

$$\mathbf{A} \leftarrow \mathbf{A}. * \mathbf{X}\mathbf{S}^T./\mathbf{A}\mathbf{S}\mathbf{S}^T \tag{25}$$

$$\mathbf{S} \leftarrow \mathbf{S}. * \mathbf{A}^T\mathbf{X}./(\mathbf{A}^T\mathbf{A}\mathbf{S} + \frac{\lambda}{2}\mathbf{S}^{q-1}) \tag{26}$$

Similar to the $L_{1/2}$ case, we can simplify the solution by defining a column-wise objective function

$$\mathcal{C}(\mathbf{s}) = \frac{1}{2}\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_q \tag{27}$$

Following the induction in Equations (14)-(20), we get

$$\mathbf{s}^{(t+1)} = \mathbf{s}^t. * \mathbf{A}^T\mathbf{x}./\left(\mathbf{A}^T\mathbf{A}\mathbf{s}^t + \tfrac{\lambda}{2}\left(\mathbf{s}^t\right)^{q-1}\right)$$

The computational complexity of the algorithm is similar to that of standard NMF, except that the computation cost of $S^{-1/2}$ is known as $O((KN)^2)$. Hence, the computational complexity of $L_{1/2}$-NMF is $O(LKN + (KN)^2)$ for each optimization iteration. The computational costs of other $L_q(0 < q < 1)$-NMF are usually larger than $L_{1/2}$-NMF, which, for large hyperspectral imagery, favours the use of $p = 1/2$ for our method.

B. Sparsity Analysis

$L_q(0 < q < 1)$ regularizer can yield sparser solutions than its L_1 counterpart, and it approximating the L_0 regularizer as q approaches 0 (see Figure 1). Unfortunately, how to determine the optimal value of q is still an open problem. Up to now, no learning or optimization methods have been reported. In practice, a sub-optimal q can be determined via an intensive search, which, despite effective, can be inefficient.

The sparsity of the $L_q(1/2 \leq q < 1)$ solution increases as q decreases, whereas the sparsity of the solution for $L_q(0 < q \leq 1/2)$ shows little change with respect to q [45], [46], making $q = 1/2$ a good regularizer choice.

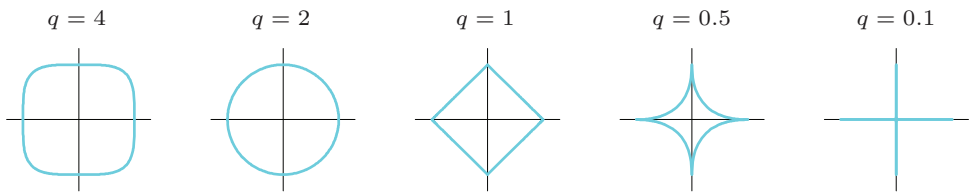


Fig. 1. The shape of L_q regularizers with different values of q.

Moreover, when the full additivity constraint $\sum_{k=1}^{K} s_k = 1$ is imposed, L_1 can not further promote sparsity. This is due to the fact that the L_1 regularizer is linked to the additivity constraint. The difference between the two lies in that the former enforces the sum of end-member abundances to be small, while the latter enforces it to be unity. Thus, the L_1 regularizer can be seen as a relaxation of the full additivity constraint, and, some cases, yields non-unique solutions [47]. For example, if we assume there exist three end-members, and the sum of abundances of a pixel is 1, for L_1 regularizer, the sparsity of $(0.3, 0.3, 0.4)$ and $(1, 0, 0)$ are the same, whereas for $L_q(0 < q < 1)$ regularizers, $(1, 0, 0)$ is sparser than $(0.3, 0.3, 0.4)$. To further illustrate the responses to the sparseness of the regularizers, in section V, we present an experiment on the synthetic data, which shows that $q = 1/2$ is a good choice for L_q regularizers with $0 < q < 1$, while L_1 may not strictly enforce sparsity.

C. Geometric interpretation

As summarized in [48], most linear spectral unmixing methods can be categorized into geometric and statistical algorithms. The geometrical methods are based on the connection between the linear spectral unmixing model and the convex geometry. In this sense, the approach proposed in this paper is a geometric method. Here, we give the geometrical interpretation for our method and show how it is related to the some other algorithms.

In general, Equation (1) is based on two main constraints:

- (A1) Nonnegative constraint: $\mathbf{A}$ and $\mathbf{S}$ are nonnegative matrix.
- (A2) Full additivity constraint: the sum of entries in each column of $\mathbf{S}$ is unity.

The first of these is met by NMF, with all the mixed or pure pixels laying on a positive simplicial cone that is given by

$$Cone(\mathbf{A}) = \left\{\mathbf{X} = \mathbf{AS} | \mathbf{S} \in \mathbb{R}_{+}^{K \times N}\right\} \tag{28}$$

If (A2) is met, all the mixed or pure pixels lie in an affine hull that is represented by

$$affine(\mathbf{A}) = \left\{\mathbf{X} = \mathbf{AS} | \mathbf{1}_N^T \mathbf{S} = 1\right\} \tag{29}$$

This implies that, if both (A1) and (A2) are satisfied, all the mixed or pure pixels lie in a convex hull (simplex) that is represented by

$$Simplex(\mathbf{A}) = \left\{\mathbf{X} = \mathbf{AS} | \mathbf{S} \in \mathbb{R}_{+}^{K \times N}, \mathbf{1}_N^T \mathbf{S} = 1\right\} \tag{30}$$

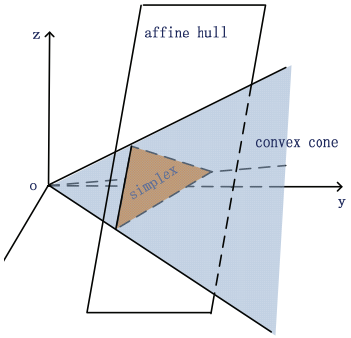
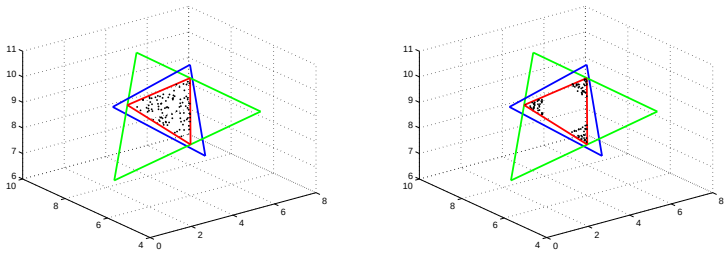


Fig. 2. Geometric interpretation of linear spectral unmixing



(a) (b)

Fig. 3. Simplex and minimim volume: (a) Uniform distribution of data points, (b) Sparse distribution of data points.

The relationship for cone, convex hull, and simplex is displayed in Figure 2. Note that, if pure pixels exist in the scene, many geometrical unmixing algorithms can be used, such as PPI, N-FINDR, and VCA. However, in the case when pure-pixels do not exist, i.e., pixels are highly mixed, the above mentioned constraints are not sufficient to produce a unique result. This can be observed in Figure 3(a), in which three simplexes meet the assumptions, but only one of them is accurate. Therefore, additional constraints have to be added to the linear spectral unmixing model, for example, those that enforce smoothness [18] on $\mathbf{A}$ and $\mathbf{S}$, or low variance on the end-member spectra [21]. In recent years, a very common practice has been to use minimum volume as an additional constraint. Methods such as ICE [32], MVES [7], MVC-NMF [43] and SISAL [9] follow Craig's unmixing criterion [6] that the end-member matrix $\mathbf{A}$ and its spanned simplex should enclose the observed pixels and minimize its volume.

The $L_{1/2}$ regularizer is closely linked to the minimum volume method through its geometric interpretation. This is as the sparsity represented by $L_{1/2}$ enforces the volume of the simplex to be minimized. This is intuitive because if the volume of simplex becomes large, the observed pixels become highly mixed (less sparse). Here, we present a simple experiment to illustrate this phenomenon. In Figure 3, we construct a simplex, which is represented by the red triangle (R) in Figure 3(a). In the figure, 100 data points spanned by three vertices are generated with uniform distribution across in this simplex. Then, we construct two more simplexes, the green (G) and blue triangles (B), which also contain all data points. Note that the volumes meet the following condition

$$V_G > V_B > V_R$$

Thus, the red simplex is the minimum one containing all the data points. The sum of the abundances of each data point in each simplex is unity. As a result, the L_1 regularizer can not distinguish the volume of the simplexes from one another. On the contrary, the $L_{1/2}$ regularizer yields a different average penalty cost of the data points for each simplex using

$$\|\mathbf{S}\|_{1/2} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1} s_n(k)^{1/2}$$

where $N = 100$ and $s_n(k)^{1/2}$ is defined in Equation (6), such that

$$\|\mathbf{S}\|_{1/2}^{(V_R)} < \|\mathbf{S}\|_{1/2}^{(V_B)} < \|\mathbf{S}\|_{1/2}^{(V_G)} \tag{31}$$

which has verified our claim that the minimum volume is related to $L_{1/2}$ regularizer.

The same observation also applies to more clustered data. If we let the 100 data points drift towards the vertices of the red simplex, as shown in Figure 3(b), then their real abundances are sparser than those corresponding to uniformly distributed data points with penalty costs fulfilling the relationship in Inequality 31.

V. Experiments

Having presented our method in the previous sections, we now turn our attention to demonstrate its utility for unmixing. Here, we employ synthetic and real-world data so as to evaluate the performance of the algorithms. The results are evaluated using Spectral Angle Distance (SAD) and Root Mean Squared Error (RMSE). The SAD is used to compare the similarity of the k^{th} end-member signature $\mathbf{A}_k$ and its estimate $\widehat{\mathbf{A}}_k$, which is defined as

$$\mathrm{SAD}_k = \arccos\left(\frac{\mathbf{A}_k^T\widehat{\mathbf{A}}_k}{\|\mathbf{A}_k\|\|\widehat{\mathbf{A}}_k\|}\right) \tag{32}$$

The RMSE is used to evaluate the the abundance estimates. It is defined as

$$\mathrm{RMSE}_k = \left(\frac{1}{N} \mid \mathbf{S}_k - \widehat{\mathbf{S}}_k \mid^2\right)^{\frac{1}{2}} \tag{33}$$

where $\widehat{\mathbf{S}}_k$ is the ground-truth abundance matrix for the k^{th} end-member.

A. Synthetic Data

Firstly, we present a quantitative analysis of our method on synthetic data. Unless specifically mentioned, the proposed $L_{1/2}$-NMF algorithm is compared against three alternatives. These are the standard NMF in [17], the L_1-NMF [40] and the L_2-NMF method [18].

For our synthetic data experiments, ten spectral signatures are chosen from the United States Geological Survey (USGS) digital spectral library [49]. Figure 4 shows six example end-member signatures used for all the following experiments. The other four spectral signatures that are not displayed in the figure include Chlorite HS179.3B,

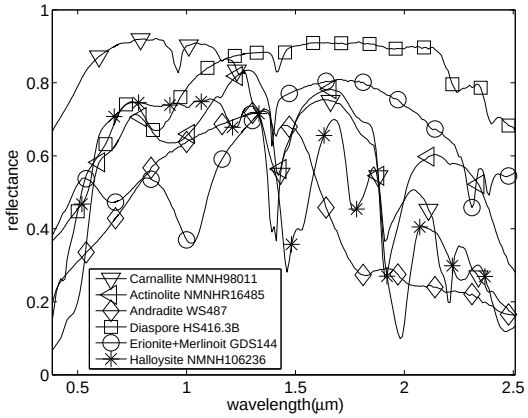


Fig. 4. Example spectral signatures from USGS used in our synthetic data experiments.

Axinite HS342.3B, Galena S26-39, and Goethite WS220, which are only used to evaluate the performance of the unmixing methods with respect to the variation of endmemeber number. To generate the synthetic data, we compute the ground-truth abundances in a similar manner to that in [43]. That is, we depart 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 mentioned above. We then apply a $(z+1) \times (z+1)$ low pass filter to 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 mixed in higher degree. If the abundance of a pixel is larger than θ, this pixel is replaced with a mixture of all end-members with equal abundances. This threshold parameter can be used to produce the synthetic data with different levels of sparseness, i.e., the smaller the θ is, the synthetic data is mixed in a higher degree, and in turn, with less sparsity. The measure of average sparseness for all the pixels in an image is defined as

$$sparse(\mathbf{A}) = \frac{1}{\sqrt{K}} \sum_k \frac{\sqrt{K} - \|\mathbf{A}_k\|_1 / \|\mathbf{A}_k\|_2}{\sqrt{K} - 1} \tag{34}$$

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

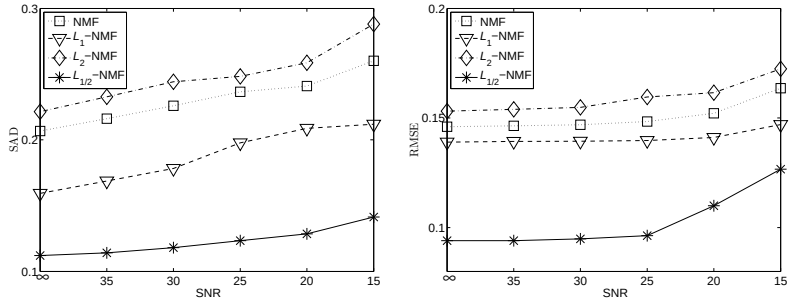
Note that, from the synthetic data generation process introduced above, its natural to perform a quantitative analysis on the four methods, i.e. our proposed $L_{1/2}$-NMF unmixing algorithm and the other three alternatives, with respect to the SNR, the sparseness, the image size and the number of end-members.

Thus, we commence by providing a sensitivity study where we have used zero-mean white Gaussian noise, which, when substituted into Equation (1), yields the following signal-to-noise ratio (SNR)

$$\text{SNR} = 10 \log_{10} \frac{E[(\mathbf{As})^T (\mathbf{As})]}{E[\mathbf{e}^T \mathbf{e}]} \tag{35}$$

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

In Figure 5 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



(a) (b)

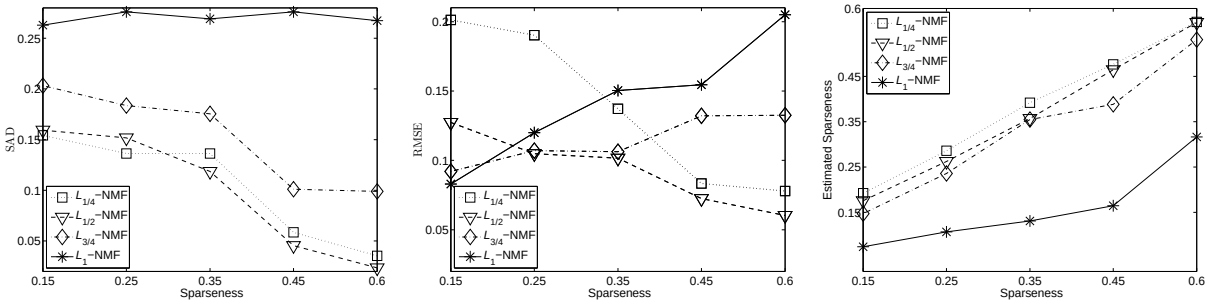
Fig. 5. Results on synthetic data: SAD (a) and RMSE (b) as a function of SNR.

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 presented here aims at evaluating the performance of the $L_q(0 < q \leq 1)$ regularizer with respect to sparsity changes. Here, L_2-NMF is removed due to its low sparse solution, while $L_{1/4}$ and $L_{3/4}$-NMFs are added for comparison. Figure 6 shows the plots of the unmixing results with different sparseness levels of the end-member abundances. Here the parameters are set as $K = 6$ and $z = 7$, where θ is used to control the sparseness levels. From figure 6(a), it can be seen that the SAD of L_1-NMF is not overly affected by the variation of the sparsity. This observation supports our claim that the L_1 regularizer cannot impose strict sparsity on the solution when the full additivity constraint is enforced. On the contrary, the SADs from the other three L_q-NMFs ($q = 1/4, 1/2, 3/4$) decrease with the increase of sparseness. This means that L_q-NMF ($0 < q < 1$) is a better option for the problem than $q = 1$.

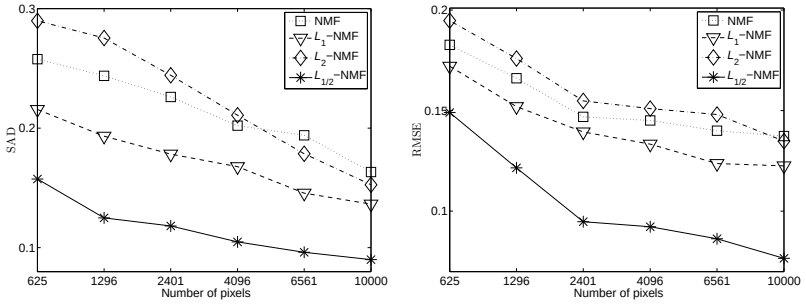
In Figure 6(b), with the increase of the sparsity, there is also an increase of the RMSE for L_1-NMF and $L_{3/4}$-NMF, while $L_{1/4}$-NMF and $L_{1/2}$ methods show clear advantages. Figure 6(c) shows the true sparseness against the estimated sparseness. It is obvious that the sparsity obtained by L_1-NMF is much lower than the true sparseness along with the increase of the abundance sparsity. Meanwhile, the estimated sparsities by L_q-NMF ($0 < q < 1$) are closer to the ground truth. Furthermore, from Figure 6, we found that the results of $L_{1/2}$-NMF and $L_{1/4}$-NMF are very similar, which validates the observation that the sparsity of the $L_q(1/2 \leq q < 1)$ increases as q decreases, whereas the sparsity of the solution for $L_q(0 < q \leq 1/2)$ does not change greatly with respect to q. This makes $q = 1/2$ a good choice when the L_q ($0 < q < 1$) regularizer is used 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 $625, 1296, \ldots, 10000$, which corresponds to $z = 5, 6, \ldots, 10$. Again, we have used six end-members and set the SNR = 30dB and $\theta = 0.7$. From Figure 7, we can see that the performances of the four methods increase as the size of the data grows. This is expected, since a large training set makes the solution space more



(a) (b) (c)

Fig. 6. 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.

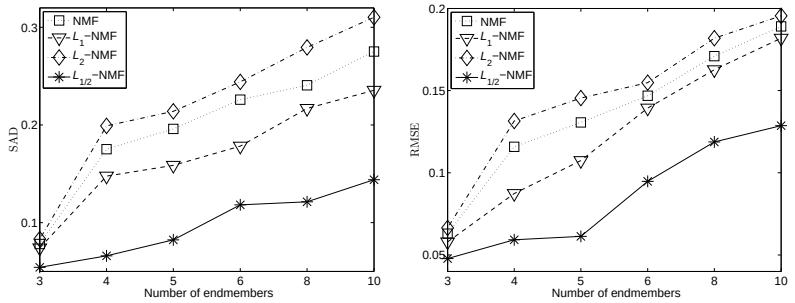


(a) (b)

Fig. 7. Results on synthetic data: SAD (a) and RMSE (b) as a function of the number of pixels in the scene.

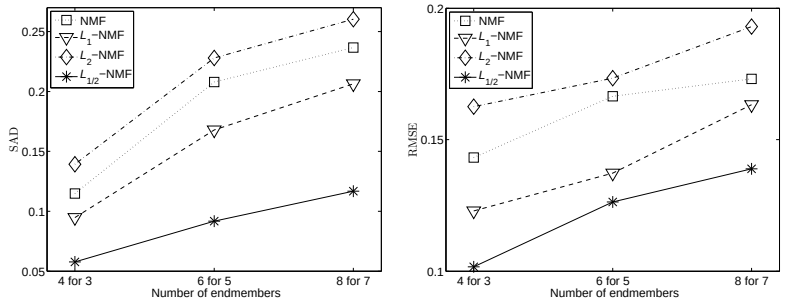
stable and constrained, which, in turn, increases the likelihood of finding an optimal or near-optimal solution.

Finally, we examine the performance of the four methods when the number of end-member changes. We also evaluate 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 8 shows the performance of the four methods when the SNR = 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. When the number of end-members increases, the differences between the algorithms become apparent, with the $L_{1/2}$-NMF consistently yielding 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. Figure 9 shows the plot of the unmixing results when the number of end-members is overestimated by 1. As before, the results obtained by our $L_{1/2}$-NMF method are more accurate than those of the other three alternatives. This is in-line with our other experiments, 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.



(a) (b)

Fig. 8. Results on synthetic data: SAD (a) and RMSE (b) as function of the number of endmembers.



(a) (b)

Fig. 9. Results on synthetic data: SAD (a) and RMSE (b) as function of the number of end-members overestimated by 1.

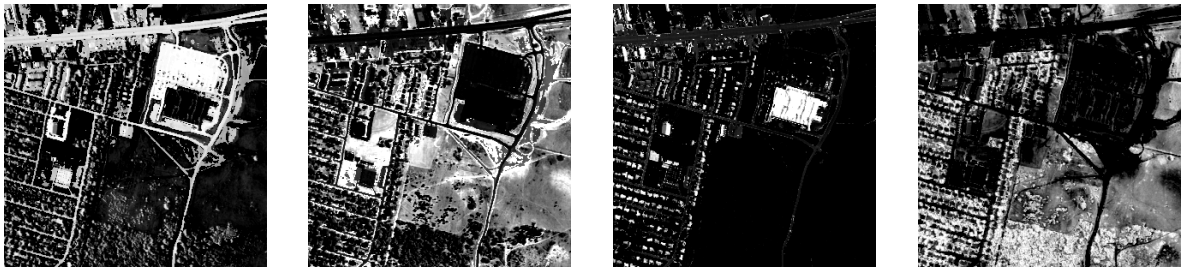
B. Experiments on Real-word Data

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 is given by the Urban HYDICE hyperspectral image. The image depicts the scene in Figure 10 and is of size 307×307 being composed of 210 spectral channels with spectral resolution of 10nm in the 400nm and 2500nm range. 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 of interest: asphalt, grass, roof and tree. Figure 11 displays the ground truth for the abundance fractions of the end-members. In these images, and from now on, the brightness of a pixel denotes the abundance of the end-member under consideration.

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, results for other state-of-the-art methods are also presented. These are VCA [3], MVC-NMF [43], PSNMFSC [20] and SISAL [9]. VCA is a popular geometric approach while MVC-NMF and PSNMFSC are both constrained NMF methods. The former adopts minimum volume as constraint while the latter 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

Fig. 10. Urban HYDICE hyperspectral dataset at band 80.



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

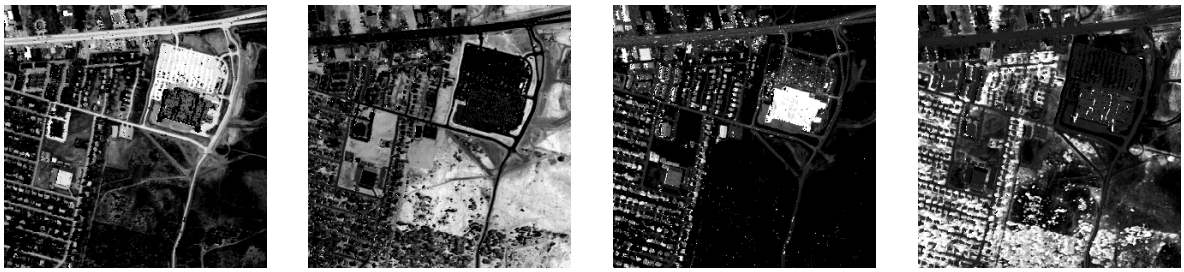
Fig. 11. Ground-truth abundance maps for four targets in the Urban HYDICE hyperspectral data.

cost point of view.

Figure 12 illustrates the separated abundance fractions for each end-member as delivered by our $L_{1/2}$-NMF. Meanwhile, Figure 13 displays the estimated $L_{1/2}$-NMF end-member signatures with respect to the USGS library spectra. The figures show that the results from $L_{1/2}$-NMF are in good accordance with the real end-member signatures and abundances. Table I gives the mean and standard variances of the SAD over 10 runs for these methods. From the table, it can be seen that, in general, the results obtained by our $L_{1/2}$-NMF are better than those yielded by the other algorithms.

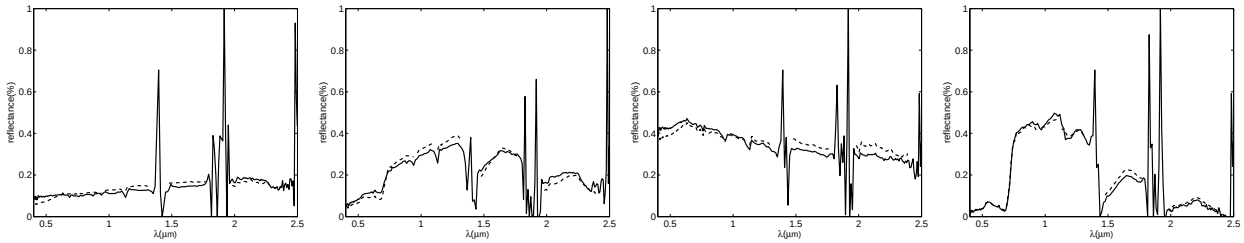
In [50], the road is further divided into asphalt and concrete, whereas the roof is divided into roof #1 and roof #2/shadow. In order to test our method under this setting, we performed experiments with six end-members. From the Figures 14, 15, 16 and Table II, we also find that the results obtained by our $L_{1/2}$-NMF are better than those yielded by the alternatives.

We now turn our attention to the second real-world data, which 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 [51]. In recent years, the Cuprite data set has been widely used for hyperspectral unmixing research [3], [43]. Figure 17 displays



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

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



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

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

the 80^{th} 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 bands.

According to [3], 14 types of minerals are presented in the scene. Its worth mentioning that variants of the same mineral with slightly different spectra can be considered as the same end-member and, hence, we set $K = 10$. In Figure 18, 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, with the smallest mean and variance.

VI. 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_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 estimates the end-member signatures and abundances using a rescaled gradient descend method. This algorithm is further extended to L_q $(0 < q < 1)$ regularizers. The sparsity property of the L_q regularizer and its intrinsic link to minimum volume

TABLE I

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

	$L_{1/2}$-NMF	L_1-NMF	MVCNMF	VCA	SISAL	PSNMFSC
Asphalt	**0.1352± 0.88**	0.3347± 0.86	0.5073± 5.65	0.3091± 1.10	0.1868± 3.71	0.1742± 3.83
Grass	**0.2097 ± 2.65**	0.2282±4.82	0.2908±2.25	0.3196±8.40	0.2598±1.06	0.2877± 5.06
Tree	**0.0658±3.05**	0.1436±0.56	0.2124±5.97	0.2110±2.57	0.1724±2.77	0.1213± 3.40
Roof	0.2441±6.03	0.4664±0.35	0.2783±5.37	0.7619±0.20	0.2097±0.52	**0.2050± 3.59**
Mean	**0.1637±3.15**	0.2932±1.65	0.3222±4.81	0.4004±3.07	0.2072±2.02	0.1971± 3.97

TABLE II

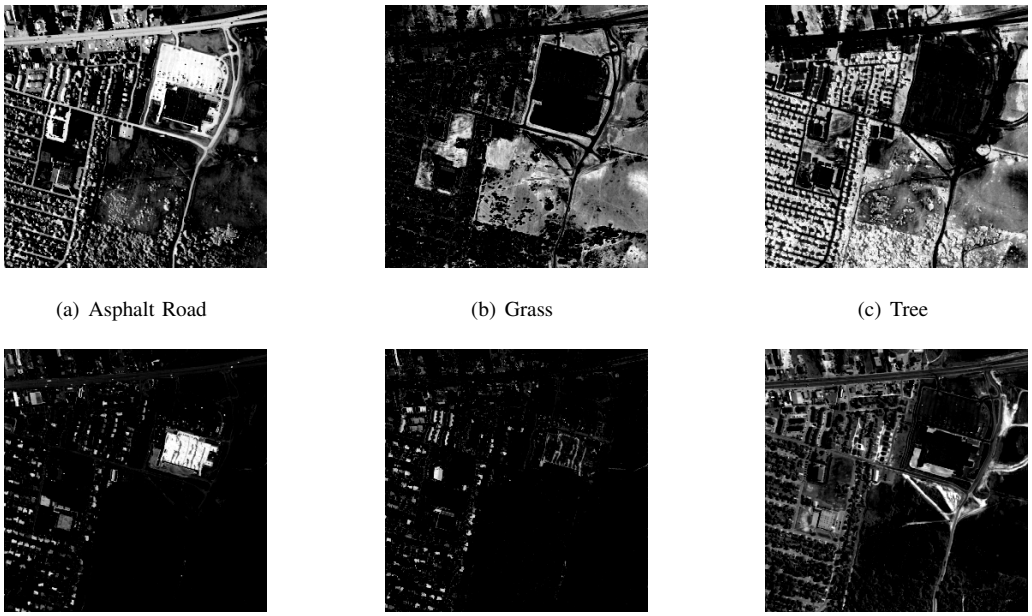
SAD AND THE STANDARD VARIANCE (%) RESULTS ON THE URBAN HYDICE DATA WITH 6 END-MEMBERS.

	$L_{1/2}$-NMF	L_1-NMF	MVCNMF	VCA	SISAL	PSNMFSC
Asphalt road	0.2302± 1.03	0.3739± 1.81	0.4978± 5.49	0.3304± 7.70	**0.2263± 2.03**	0.5151± 5.45
Grass	0.2692 ± 2.95	0.3772±3.40	0.3334±4.57	0.5104±3.06	0.3655±3.29	**0.1781± 6.62**
Tree	**0.0414±1.14**	0.1523±1.45	0.1727±4.99	0.3108±3.79	0.2513±1.59	0.1717± 8.12
Roof #1	**0.1000±3.56**	0.7134±2.66	0.2574±4.35	0.9529±6.56	0.4339±5.53	0.2624± 5.70
Roof #2/shadow	**0.2617±4.27**	0.5849±0.97	0.4389±5.06	0.5409±3.22	0.3373±4.65	0.4477± 6.65
Concrete road	**0.1570±0.83**	0.5136±0.60	0.2882±4.34	0.3721±0.46	0.3119±8.84	0.3226± 8.65
Mean	**0.1766±2.30**	0.4526±1.82	0.3314±4.80	0.5029±4.13	0.3210±4.32	0.3163± 6.87

TABLE III

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

	$L_{1/2}$-NMF	L_1-NMF	MVCNMF	VCA	SISAL	PSNMFSC
Alunite	0.1660± 0.61	0.2874± 6.83	0.1421± 5.92	0.1016± 3.19	0.2103± 5.63	**0.0979± 2.91**
Andradite	**0.0545 ± 0.14**	0.2380±0.87	0.1488±5.11	0.0715±3.18	0.0869±4.57	0.0938± 11.04
Buddingtonite	0.1626±0.25	0.3448±8.86	0.2058±4.44	**0.1317±1.99**	0.1413±4.98	0.3303± 8.62
Dumotierite	**0.0918±1.27**	0.2112±4.87	0.1553±3.90	0.0988±1.30	0.1221±4.77	0.2310± 2.30
Kaolinite	**0.1441±4.46**	0.3793±3.46	0.1979±4.15	0.2418±4.85	0.2087±3.97	0.3493± 8.54
Montmorillonite	**0.1206±2.26**	0.3014±4.75	0.1339±9.87	0.1357±4.52	0.1605±4.03	0.2591± 6.95
Muscovite	0.1292±2.39	0.3801±5.01	**0.0914±9.28**	0.1349±4.01	0.1398±4.58	0.2275± 7.99
Nontronite	**0.0850±0.08**	0.2292±0.92	0.1779±8.03	0.0920±1.81	0.0887±3.89	0.1379± 6.82
Pyrope	**0.0596±0.39**	0.3188±0.43	0.2078±7.71	0.1257±3.77	0.1482±3.96	0.1430± 10.23
Sphene	0.1002±0.28	0.2718±5.23	0.1439±3.73	0.0794±5.16	**0.0713±2.26**	0.1134± 3.04
Mean	**0.1114±1.21**	0.2962±4.12	0.1585±6.21	0.1213±3.38	0.1378±4.26	0.1983± 6.84



(a) Asphalt Road (b) Grass (c) Tree

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

Fig. 14. Ground-truth abundance maps for six targets in the Urban HYDICE hyperspectral data.

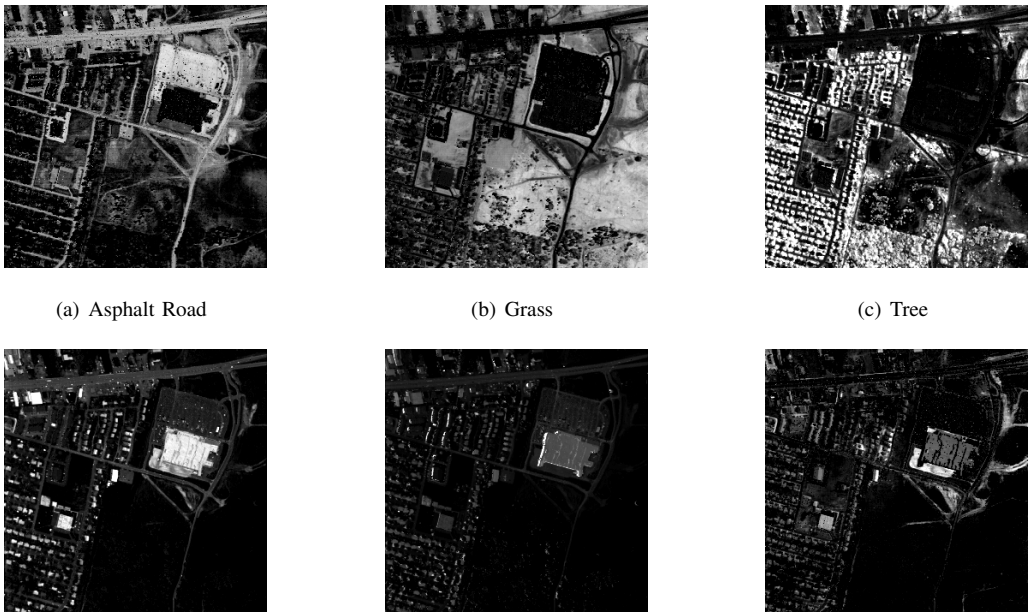
methods have been discussed.

We have illustrated the advantages of our unmixing method on synthetic and real-world data and compared our method against 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 emphasize that the $L_{1/2}$-NMF method presented here is quite general in nature and can be readily applied to other settings in which nonnegative sparse matrix factorization is a valuable computational tool. Furthermore, the method presented here can easily incorporate constraints found elsewhere in the literature.

Apart from the multiplicative iterative algorithm proposed in this paper, projected gradient algorithm, alternating least squares algorithm, and quasi-Newton algorithms are also widely used for NMF and related problems. Most of these algorithms can be extended to $L_{1/2}$-NMF in modified forms. New optimization methods are worth exploring so as to extend the $L_{1/2}$-NMF to wavelet-subspaces, Markovian formulations based upon spatial consistency constraints and 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.

VII. ACKNOWLEDGMENT

The authors would like to thank the reviewers for the many valuable comments and suggestions. This work was supported by the China-Australia Special Fund for Science and Technology Cooperation No.61011120054, National



(a) Asphalt Road (b) Grass (c) Tree

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

Fig. 15. Urban HYDICE results: Abundance maps estimated using $L_{1/2}$-NMF for the six targets.

Natural Science Foundation of China No.60902070 and No.60872071 and the Australian Government as represented by the Department of Innovation, Industry, Science and Research under the International Science Linkages (ISL) program, Round 9 of the Australia-China Special Fund for S&T (ISL reference number CH090174).

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.

[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.

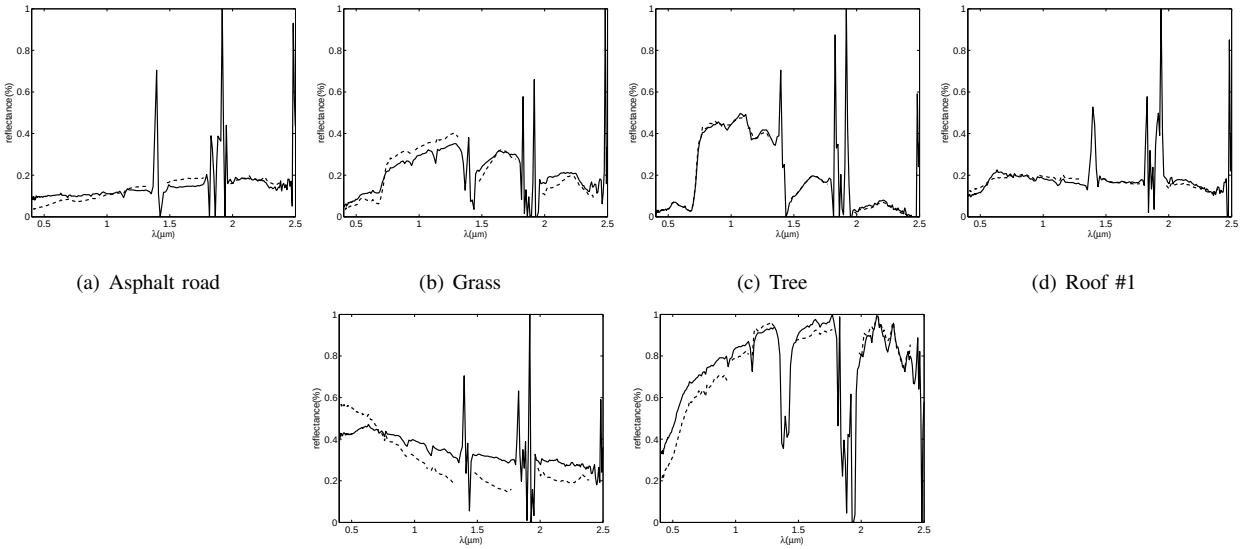
[5] A. Zymnis, S. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Proceedings of the 41st Asilimar Conference on Signals, Systems, and Computers*, 2007, pp. 1164–168.

[6] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 1, pp. 99–109, Jan. 1994.

[7] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proceedings of IGARSS*, 2008, pp. 4418–4432.

[8] 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.

[9] J. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. of the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2009, pp. 1–4.



(a) Asphalt road (b) Grass (c) Tree (d) Roof #1

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

Fig. 16. Urban HYDICE results: Comparison of the USGS library spectra (solid line) with the signatures extracted by $L_{1/2}$-NMF (dotted line) with 6 end-members.

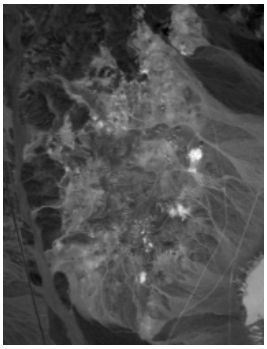


Fig. 17. The AVIRIS Cuprite image (band 80).

[10] A. Ambikapathi, T. Chan, W. Ma, and C. Chi, "A robust alternating volume maximization algorithm for endmember extraction in hyperspectral images," in *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2010, pp. 1–4.

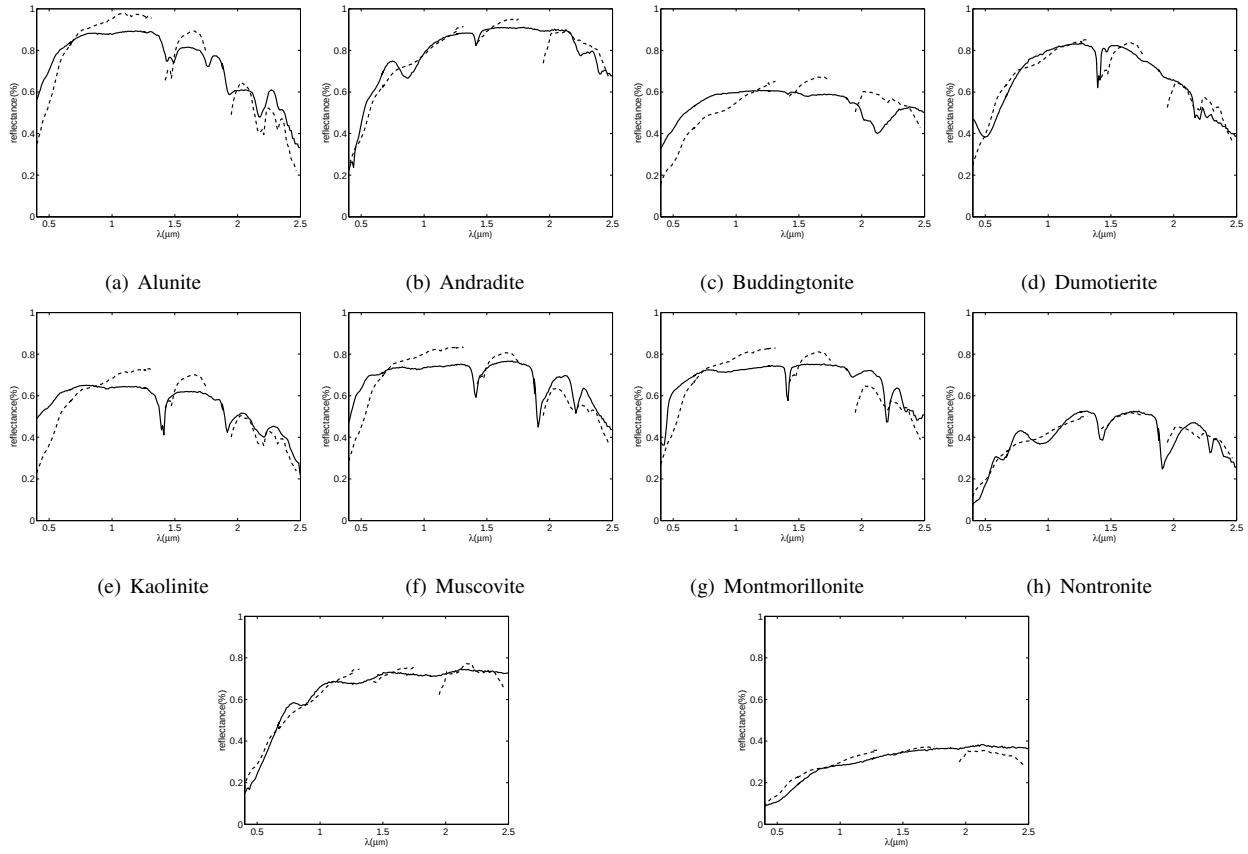
[11] J. Chen, X. Jia, W. Yang, and B. Matsushita, "Generalization of subpixel analysis for hyperspectral data with flexibility in spectral similarity measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2165–2171, 2009.

[12] 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.

[13] 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.

[14] 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.

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



(a) Alunite (b) Andradite (c) Buddingtonite (d) Dumotierite

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

(i) Pyrope (j) Sphene

Fig. 18. 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).

York, 2003.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2590–2602, 2010.

[22] D.Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *Proc. NIPS16*, 2004.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] M. D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *to appear in IEEE Transactions on Geoscience and Remote Sensing*, 2011.

[30] 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.

[31] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 446–450, 2007.

[32] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[33] 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.

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

[35] 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.

[36] 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.

[37] 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.

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

[39] D. R. Hunter and K. Lange, "A tutorial on mm algorithms," *The American Statistician*, vol. 58, pp. 30–37, 2004.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

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

[45] Z. Xu, "Data modeling: Visual psychology approach and $l_{1/2}$ regularization theory," in *Proceeding of the International Congress of Mathematicians*, 2010.

[46] Z. Xu, H. Guo, Y. Wang, and H. Zhang, "The representative of $l_{1/2}$ regularization among $l_q(0 < q < 1)$ regularization: An experimental study based on a phase diagram," in *Submitted*, http://gr.xjtu.edu.cn:8080/web/zbxu.

[47] A. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse s olutions to underdetermined systems of equations," *IEEE Trans. on Information Theory*, vol. 54, no. 11, pp. 4813–4820, 2008.

[48] J. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing: Geometrical, statistical and sparse regression-based approaches." in *SPIE Remote Sensing Europe, Image and Signal Processing for Remote Sensing Conference*, 2010.

[49] 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.

[50] X. Liu, W. Xie, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *to appear in IEEE Transactions on Geoscience and Remote Sensing*, 2011.

[51] 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.