# Sparse Hyperspectral Unmixing Using An Approximate $L_0$ Norm☆

Wei Tang, Zhenwei Shi*, Durengjan Borjigen

*Image Processing Center, School of Astronautics, Beihang University, Beijing, P.R. China, 100191*

**Abstract**

Sparse unmixing aims at finding an optimal subset of spectral signatures in a large spectral library to effectively model each pixel in the hyperspectral image and compute their fractional abundances. In most previous work concerned with the sparse unmixing, $L_2$ norm is used to measure the error tolerance and the $L_1$ norm is added as the sparsity regularization. However, in some applications, using $L_1$ norm to measure the error tolerance has significant robustness advantages over the $L_2$ norm. Besides, in some cases, using a smooth function to approximate the $L_0$ norm can obtain more accurate results than the $L_1$ norm in the field of sparse regression. Thus, in this paper, we consider the two alternative choices for sparse unmixing. A reweighted iteration algorithm is also proposed so that the unconvex regularizer (smoothed $L_0$ norm) can be efficiently solved through transforming it into a series of weighted $L_1$ regularizer problems. Experimental results on both synthetic and real hyperspectral data demonstrate the efficacy of the new models.

☆The work was supported by the National Natural Science Foundation of China under the Grants 61273245 and 91120301, the 973 Program under the Grant 2010CB327904, the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No. BUAA-VR-12KF-07), and Program for New Century Excellent Talents in University of Ministry of Education of China under the Grant NCET-11-0775 and the Beijing Natural Science Foundation (Non-negative Component Analysis for Hyperspectral Imagery Unmixing) under the Grant 4112036.The work was also supported by Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, P.R. China.

*Corresponding author. Tel.: +86-10-823-39-520; Fax: +86-10-823-38-798.

*Email address:* `shizhenwei@buaa.edu.cn` (Zhenwei Shi)

*Keywords:* Hyperspectral Unmixing, Sparse Hyperspectral Unmixing, Spectral Library, $L_0$ Minimization, Reweighted $L_1$ Minimization

## 1. Introduction

Among the remote sensing modalities, hyperspectral imaging is a crucial technique that can identify the materials and their compositions in an area by making use of the spectral diversity of the different materials. Utilizing the hyperspectral remote sensing technology, we can capture hyperspectral images ranging from ultraviolet region to infrared band. As a result, we can obtain a hyperspectral data cube which is a stack of images. Each pixel of the hyperspectral imagery is represented by a spectral signature. As the hyperspectral data contain hundreds of contiguous narrow spectral band images, they have great value of applications such as mineral identification [1], Space Object Identification (SOI) [2] and etc [3]. However, due to the low spatial resolution of a sensor and the combination of distinct materials into a homogeneous mixture, each pixel in the hyperspectral imagery often contains more than one pure substance. Spectral unmixing which means to extract the pure materials (endmembers) from the spectrum mixture and estimate their corresponding fractions (abundances) is important to numerous tactical scenarios, especially when the subpixel information is valuable.

To deal with the unmixing problem, linear spectral mixture analysis techniques first identify a collection of pure constituent spectra, then express the measured spectrum of each mixed pixel as a linear combination of endmembers weighted by their fractional abundances [4]. Spectral unmixing includes two main steps. The first step, called endmember extraction, aims at extracting endmembers from the hyperspectral image. This step can be achieved in a supervised manner when prior information about the image is available. For example, one can recognize classes of pure materials in the image and select the associated endmembers to create a learning set containing samples belonging to different classes. However, sometimes this information is not available, so an automatic endmember extraction algorithm (EEA) is necessary. There are many automatic endmember extraction algorithms and we can categorize them into two types. Type one is the pure pixel based algorithms, such as the N-FINDR [5], the pixel purity index (PPI) [6] and the vertex component analysis (VCA) [7]. They assume that the data set contains at least one pure pixel for each endmember, and then extract the purest pixels from the image. Type two is minimum volume based algorithms, such as the minimum-volume enclosing simplex (MVES) [8], the minimum volume simplex analysis (MVSA) [9], the iterated constrained endmembers (ICE) [10]. The second step, called inversion, consists of computing the corresponding abundances under the constraints of nonnegativity and sum-to-one [11]. Many different algorithms have been proposed in the literature to estimate the abundances for the linear mixing model [12]. These algorithms are based on the least square principle [13], maximum likelihood estimation [14], or Bayesian methods [15, 16]. Besides the algorithms mentioned above, there are also some algorithms which can pick up the endmembers and compute the fractional abundances in one stage. One of them is the nonnegative matrix factorization (NMF) [17, 18] based algorithms. Such algorithms including minimum volume constrained nonnegative matrix factorization (MVC-NMF) [19], $L_{1/2}$ sparsity-constrained nonnegative matrix factorization ($L_{1/2}$-NMF) [20] and manifold regularized sparse NMF [21] can factorize the observed data matrix into the endmember matrix and fractional matrix. However, they need to know the number of endmembers in advance and identify materials by comparing them to the pure spectral signatures in the spectral library.

In recent years, as the spectral libraries become available, hyperspectral unmixing problem can be solved in a semi-supervised fashion. The endmembers can be derived from a spectral library (potentially very large) and used for unmixing [22]. As the libraries are usually very large, this approach generally results in a sparse solution. Consequently, we call this approach sparse unmixing. Several sparse regression techniques have been used for sparse unmixing in [22], including orthogonal matching pursuit (OMP), basis pursuit (BP), BP denoising and iterative spectral mixture analysis (ISMA). Recently, some sparse unmixing methods have been proposed that exploit the contextual information [23, 24] and subspace nature [25] of the hyperspectral images to obtain a much better result. As the approach of sparse unmixing takes advantage of the increasing availability of the spectral libraries, the abundance estimation process no longer depends on the availability of pure spectral signatures nor on certain endmember extraction algorithms to indentify such pure signatures.

To the best of our knowledge, in most previous work concerned with sparse unmixing, the $L_2$ norm is used to measure the error tolerance because its computational simplicity compared with the $L_1$ norm, and the convex relaxation methods (namely adding an $L_1$ regularization) are often considered as it can convert the untractable $L_0$ norm problem into a more tractable one. However, using the $L_1$ norm to measure the error tolerance are known to have significant robustness advantages over the conventional $L_2$ norm in many applications [26, 27]. Besides, using a smoothed $L_0$ norm to replace the $L_0$ norm can provide better accuracy than the $L_1$ norm in some cases [28]. Thus, in this paper, we consider exploring whether the two alternative modifications can get better sparse unmixing results. Nonetheless, the smoothed $L_0$ norm is not a convex function. Thus, a reweighted iteration algorithm is proposed so that the unconvex regularizer can be efficiently solved through transforming it into a series of weighted $L_1$ regularizer problems.

This paper is organized as follows. In Section 2 we introduce the sparse hyperspectral unmixing model. Section 3 presents the unmixing algorithms considered in our study. Section 4 conducts the experiments to compare the performances of different algorithms using both the synthetic data and real data. Finally Section 5 concludes with some remarks.

## 2. Sparse Hyperspectral Unmixing Model

Spectral unmixing aims at estimating the fractional abundances of endmembers in each mixed pixel. The linear mixing model assumes that the observed spectrum of a pixel can be expressed as a linear combination of the spectra of the endmembers. Assume $y = (y_1, \dots, y_l)^T$ is the observed hyperspectral spectrum, where $l$ is the number of the spectral bands. The $i$-th element of $y$ represents the reflectance of the observed spectrum at band $i$. The linear mixing model can be expressed mathematically as follows:

$$y_i = \sum_{j=1}^{q} M_{ij} v_j + n_i \tag{1}$$

where $q$ is the total number of endmembers in the hyperspectral data. $M_{ij}$ is the reflectance of the $j$-th endmember at band $i$. $v_j$ is the fractional abundance of the $j$-th endmember for the observed spectrum, and $n_i$ represents the error term for the observed pixel at spectral band $i$. Eq. (1) can be rewritten in a compact matrix form as:

$$y = Mv + n \tag{2}$$

where $M$ is an $l$-by-$q$ matrix and each column of $M$ represents the reflectance of an endmember, $v$ is a $q$-D vector containing the respective fractional abundances of the endmembers and $n$ is an $l$-D vector which represents the error term.

In the sparse unmixing model, endmembers are involved in a large spectral library (denoted by $A$). Suppose there are $p$ different spectra contained in the spectral library, then Eq. (2) can be written as follows:

$$y = Ax + n \tag{3}$$

where $A$ is an $l$-by-$p$ matrix and $x$ is a $p$-D vector each element of which corresponds to the fractional abundance of a signature in the spectral library for the observed pixel. Because $q$ is much smaller than $p$, the vector of fractional abundances $x$ is sparse. Fig. 1 shows the concept of sparse unmixing. Due to the physical constraint, the fractional abundances of the endmembers cannot be negative and their sum should be one. These constraints are known as the sum-to-one and the nonnegativity constraints:

$$\sum_{i=1}^{p} x_i = 1 \quad \text{(sum-to-one)} \tag{4}$$

$$x_i \geq 0 \quad \text{(nonnegativity)} \tag{5}$$

where $x_i$ is the $i$-th element of $x$.

We first do not consider the two constraints in Eq. (4) and Eq. (5). To find sparse solutions to the unmixing problem, we should solve the following optimization problem:

$$\min_x \|x\|_0 \quad \textit{subject to} \quad y = Ax \tag{6}$$

where $\|x\|_0$ is the number of non-zero entries in $x$. Because of noise and modeling errors, most previous work about sparse unmixing replaced Eq. (6) with the following optimization problem:

$$\min_x \|x\|_0 \quad \textit{subject to} \quad \|y - Ax\|_2 < \epsilon \tag{7}$$

where $\epsilon$ is a preset threshold corresponding to the noise and modeling errors. However, $L_1$ methods are known to have significant robustness advantages over $L_2$ methods in many applications [26, 27]. Thus in this paper, we consider using the $L_1$ norm to replace the $L_2$ norm in Eq. (7)

$$\min_x \|x\|_0 \quad \textit{subject to} \quad \|y - Ax\|_1 < \epsilon \tag{8}$$

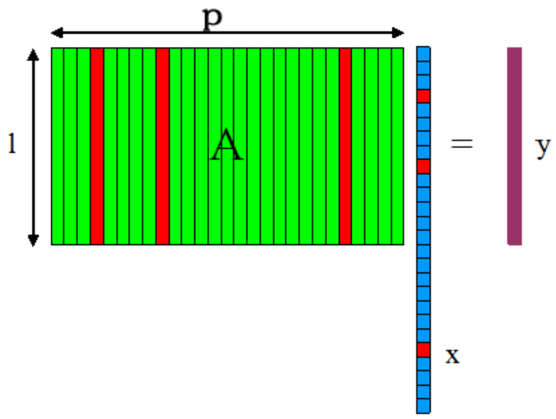


Figure 1: Concept of the sparse hyperspectral unmixing.

The unconstrained versions of Eq.(7) and Eq.(8) are as follows:

$$\min_x \|y - Ax\|_2^2 + \lambda \|x\|_0 \tag{9}$$

$$\min_x \|y - Ax\|_1 + \lambda \|x\|_0 \tag{10}$$

where $\lambda > 0$ is the regularization parameter. Theoretically, a bigger $\lambda$ leads to a sparser solution for the two models in Eq. (9) and Eq. (10). The relationship between $\lambda$ and $\epsilon$ is that $\lambda \to 0$ when $\epsilon \to 0$. In Eq. (9), when we choose the parameter $\lambda = 0$, it becomes the well-known least square estimation. The least squares estimator has the smallest mean squared error of all linear estimators with no bias [29].

We call Eq. (9) and Eq. (10) as $L_2 - L_0$ model and $L_1 - L_0$ model, respectively. The optimization problems presented in Eq. (7-10) are NP-hard and it is difficult to solve them in a direct way. Therefore we consider some algorithms which can get an approximate solution, such as greedy algorithms [30] and convex relaxation. There are many greedy algorithms that can compute a sparse solution, such as the classical pursuit algorithm Orthogonal Matching Pursuit (OMP) [31], Simultaneous Orthogonal Matching Pursuit (S-OMP) [32, 33] and Subspace Pursuit (SP) [34]. Convex relaxation method such as Basic Pursuit [35] use $L_1$ norm to replace the $L_0$ norm in Eq. (7-10).

By replacing the $L_0$ norm in Eq. (9) and Eq. (10) with $L_1$ norm, we can have optimization problems as follows:

$$\min_x \|y - Ax\|_2^2 + \lambda\|x\|_1 \tag{11}$$

$$\min_x \|y - Ax\|_1 + \lambda\|x\|_1 \tag{12}$$

We call them $L_2 - L_1$ model and $L_1 - L_1$ model, respectively.

## 3. Algorithms for Sparse Unmixing

### *3.1. An Approximative Approach With Smoothed $L_0$ Norm*

As $L_0$ norm is not continuous, it is not easy to solve the problems in Eq. (9) and Eq. (10) directly. In this paper, we consider using a continuous function to approximate $L_0$ norm, which has three advantages. Firstly, such a representation provides a smooth measure of sparsity. Secondly, such a definition of sparsity could tolerate noise to some extend [28]. Thirdly, the smoothed $L_0$ norm can provide better accuracy than the $L_1$ norm in some applications [36]. To obtain these benefits, functions must have the following properties: $f(0) = 0$ and $f(1) = 1$.

In this paper, we propose a new function to approximate the $L_0$ norm. For the $i$-th element in $x$, the proposed function is defined as:

$$f(a, x_i) = \frac{1}{\log_a(ax_i)} \quad (0 < a < 1, 0 \leq x_i \leq 1) \tag{13}$$

where $a$ is a given scaler. Then the smoothed $L_0$ norm of vector $x$ is defined as

$$\begin{aligned} F(a, x) &= \textstyle\sum_{i=1}^{p} f(a, x_i) \quad (14)\\ &= \textstyle\sum_{i=1}^{p} \frac{1}{\log_a(ax_i)} \quad (15) \end{aligned}$$

The following theorem shows how the proposed function approximates the $L_0$ norm.

**Theorem 1.** $F(a, x)$ converges to the $L_0$ norm of vector $x$ as $a \to 0$, where $x = (x_1, \ldots, x_p)^T$.

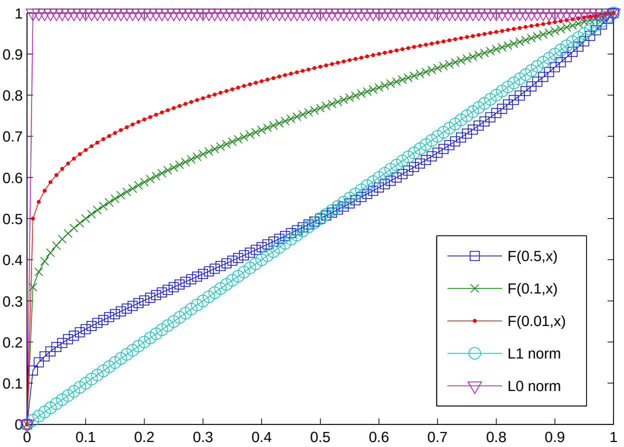


Figure 2: The graph of function $F(a, x)$ with different parameter $a$.

*Proof.* From Eq. (13), we can get

$$\begin{aligned}\lim_{a\to 0} f(a, x_i) &= \lim_{a\to 0} \frac{1}{\log_a(ax_i)} \\ &= \lim_{a\to 0} \frac{1}{1+\log_a(x_i)} \\ &= \lim_{a\to 0} \frac{1}{1+\frac{1}{\log_{x_i}(a)}} \\ &= 1 \quad \forall x_i \in (0, 1] \end{aligned} \tag{16}$$

$$f(a, 0) = 0 \quad \forall a \in (0, 1) \tag{17}$$

Combining Eq. (16) and Eq. (17), we obtain

$$\lim_{a\to 0} \sum_{i=1}^{p} f(a, x_i) = \|x\|_0 \tag{18}$$

Thus, the theorem is proved. □

Fig. 2 shows the sparsity property of the our function $F(a, x) = \sum_{i=1}^{p} f(a, x_i)$ from the aspect of geometry for $p = 1$. From Fig. 2 and theorem 1, we can find that the smaller value of $a$, the closer behavior of $F(a, x)$ to $L_0$ norm;

and the larger value of $a$, the smoother $F(a,x)$ (but worse approximation to $L_0$ norm).

Using the function $\sum_{i=1}^{p} f(a,x_i)$ to replace the $L_0$ norm, we can get new versions of Eq. (9) and Eq. (10), they can be expressed mathematically as follows:

$$\min_x \|y - Ax\|_2^2 + \lambda \sum_{i=1}^{p} f(a,x_i) \tag{19}$$

$$\min_x \|y - Ax\|_1 + \lambda \sum_{i=1}^{p} f(a,x_i) \tag{20}$$

We call the models in Eq. (19) and Eq. (20) as $L_2 - SL_0$ model and $L_1 - SL_0$ model, respectively.

### *3.2. Algorithms For Minimizing Approximative Approach With smoothed $L_0$ Norm*

As the optimization problems in Eq. (9) and Eq. (10) are NP-hard, we use Eq. (19) and Eq. (20) to replace them. Note that when $0 < a < e^{-2}$, the smoothed $L_0$ norm is not convex. And to make the behavior of the smoothed function close to the $L_0$ norm, we usually need a very small $a$ (we choose $a = 10^{-5}$ in this paper). In this case, the problems in Eq. (19) and Eq. (20) are not convex optimization problems. With the development of the non-convex optimization techniques, some effective iterative weighted algorithms have been proposed [37, 38, 39]. We use the first-order Taylor expansion of $f(a,x_i)$ at the $t$-th iteration point $x_i^t$ to replace the $f(a,x_i)$, and thus convex optimization problems can be obtained:

$$f(a,x_i) \approx f(a,x_i^t) + \frac{-\log_a e}{x_i^t \log_a^2(ax_i^t)}(x_i - x_i^t) \tag{21}$$

So Eq. (19) and Eq. (20) can be rewritten as follows:

$$\min_x \|y - Ax\|_2^2 + \lambda \sum_{i=1}^{p} \frac{-\log_a e}{x_i^t \log_a^2(ax_i^t)} x_i \tag{22}$$

$$\min_x \|y - Ax\|_1 + \lambda \sum_{i=1}^{p} \frac{-\log_a e}{x_i^t \log_a^2(ax_i^t)} x_i \tag{23}$$

Now, we present an iterative weighted algorithm to solve the problems in Eq. (22) and Eq. (23). We show that the problems of Eq. (22) and Eq. (23) can be transformed into linear programming problem and quadratic programming problem respectively. For simplicity, in the process of solving the optimization problems in Eq. (22) and Eq. (23), we use the following equation to represent the two problems:

$$\min_x P(x) + \lambda Q(x,t) \tag{24}$$

where $P(x)$ and $Q(x)$ corresponds with the first term and the second term in Eq. (22) or Eq. (23), respectively. The algorithm is shown in Algorithm 1.

---

**Algorithm 1** the iterative weighted algorithm for sparse hyperspectral unmixing.

---

1: **Initialization and parameters setting**
 iteration index: $t = 0$
 initialize $x$: $x^0 = arg\min_x P(x)$
 set the maximum iteration step $N$ and the error tolerance $\epsilon$.

2: **Main iteration**
 update $x$: $x^{t+1} \longleftarrow \arg\min_x P(x) + \lambda Q(x,t)$
 update iteration: $t \longleftarrow t+1$

**stop main iteration** if $t \geq N$ or $\frac{\|x^t - x^{t-1}\|_2}{\|x^t\|_2} < \epsilon$ is satisfied.

---

When $P(x) = \|y - Ax\|_2^2$ the optimization problem in step 2 is a quadratic programming problem:

$$\begin{aligned} x^{t+1} &= \arg\min_x P(x) + \lambda Q(x,t) \\ &= \arg\min_x x^T A^T A x + (\lambda c^T - 2y^T A)x + y^T y \end{aligned} \tag{25}$$

where $c = (c_1, \ldots, c_p)^T$ and $c_i = \frac{-\log_a e}{x_i^t \log_a^2 (a x_i^t)} x_i$. There are many methods that can solve the problem in Eq. (25), such as active set method. In matlab we can use the function 'quadprog' to solve it. Similarly, the $L_2 - L_1$ problem in Eq. (11) can also be solved by the quadratic programming.

When $P(x) = \|y - Ax\|_1$, we define vector $s^+$ and $s^-$ as follows:

$$s^+ = \frac{|y - Ax| + (y - Ax)}{2} \tag{26}$$

$$s^- = \frac{|y - Ax| - (y - Ax)}{2} \tag{27}$$

Then the optimization problem in step 2 becomes a linear programming problem:

$$
\begin{aligned}
x^{t+1} &= \arg\min s^{+} + s^{-} + \lambda c^{T} x \\
subject \quad to \quad & s^{+} - s^{-} + Ax = y
\end{aligned} \tag{28}
$$

We can use inner point method to solve it. In matlab we can use the function 'linprog' to solve it. Similarly, the $L_1 - L_1$ problem in Eq. (12) can also be solved by the linear programming.

Note that it is easy to add the nonnegativity constraint in Eq. (5) into the four models and use the linear programming or quadratic programming to solve them (we can also use the 'linprog' function or 'quadprog' function in matlab to solve them). We do not explicitly add the sum-to-one constraint in Eq. (4) all along because (i) the sum-to-one constraint should be replaced with a so-called generalized sum-to-one constraint as there is a strong signature variability in a real image, and (ii) the nonnegativity of the sources automatically imposes a sum-to-one generalized constraint [22].

## 4. Numerical Experiments on Hyperspectral Data

Having presented our method in the previous section, we now turn our attention to demonstrate its utility for sparse unmixing. The proposed $L_2-SL_0$ model and $L_1-SL_0$ model are compared with the $L_2-L_1$ model and $L_1-L_1$ model. All the considered models take into the nonnegativity constraint. Here, we employ synthetic data and real-world data in order to evaluate the performances of the algorithms. The root mean square error (rmse) [4] is used to evaluate the abundance estimations. For the $i$-th endmember, rmse is defined as

$$
\mathrm{RMSE}_i = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(X_{ij} - \hat{X}_{ij})^2}. \tag{29}
$$

Here, $X$ denotes the true abundances matrix and $\hat{X}$ represents the estimated one each column of which corresponds with the abundances of a pixel. The mean value of all the endmembers' rmses will be computed. Generally speaking, the smaller the rmse is, the more the estimation approximates the truth. The stop conditions for the four models are the same: the error tolerance $\epsilon$ is set to $10^{-3}$ and the maximum iteration number is set to 20 which is enough to guarantee the convergence.

### 4.1. Experiments with Synthetic Data

Firstly, we test the four models on synthetic data. The spectral library used in our synthetic experiments is the United States Geological Survey (USGS) [40] digital spectral library. The reflectance values of 498 materials are measured for 224 spectral bands distributed uniformly in the interval $0.4 - 2.5\ \mu m$. We choose eight spectral signatures from the USGS digital spectral library. They are Rhodochrosite HS67, Axinite HS342.3B, Chrysocolla HS297.3B, Niter GDS43 (K-Saltpeter), Anthophyllite HS286.3B, Neodymium_Oxide GDS34, Monazite HS255.3B and Samarium_Oxide GDS36. Their spectral signatures are displayed in Fig. 3.

The synthetic data are created as follows:

1) Divide the scene, whose size is $z^2 \times z^2$ ($z = 8$), into $z \times z$ regions. Initialize each region with the same type of ground cover, randomly selected from the endmember class. The endmember number is $q$ ($q = 8$). The size of spectral signature matrix $M$ is $l \times q$ ($l = 224$).

2) Generate mixed pixels through a simple $(z + 1) \times (z + 1)$ spatial low-pass filter.

3) Replace all the pixels in which the abundance of a single endmember is larger than 70% with a mixture made up of only two endmembers (the abundances of the two endmembers both equal 50%) so as to further remove pure pixels and represent the sparseness of abundances at the same time; After these three steps, we obtain the distribution of $q$ endmembers in the scene and specific values are stored in $V$ with a size of $q \times m$ ($m = z^2 \times z^2$).

4) Use linear spectral mixing model $Y = M \times V$ to generate hyperspectral data and add 30dB Gaussian white noise at the same time. The size of hyperspectral data $Y$ is $l \times m$.

The final size of the synthetic hyperspectral data is $64 \times 64 \times 224$. Note that there is no pure pixel in the synthetic hyperspectral data. All the parameters of the four algorithms are tuned to the best performances. Specifically, the $\lambda$'s in the $L_2 - SL_0$ model, $L_2 - L_1$ model, $L_1 - SL_0$ model and $L_2 - SL_0$ model are set to be 0.1, 1, 0.2 and 1, respectively. As mentioned before, the $a$'s in the $L_2 - SL_0$ model and $L_1 - SL_0$ model are set to be $10^{-5}$. In the $L_2 - SL_0$ model and $L_2 - L_1$ model, we set the initial solution

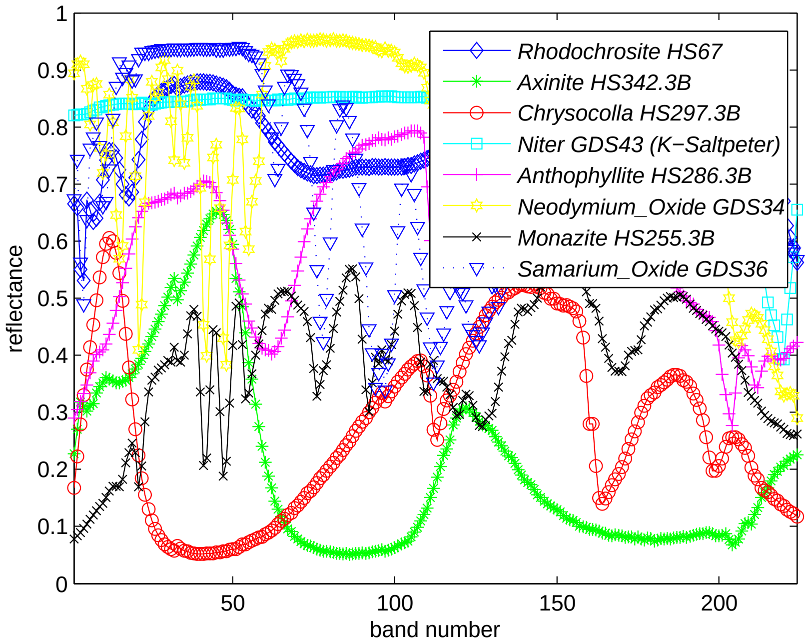


Figure 3: The eight spectral signatures used in our synthetic data experiment.

Table 1: Results obtained by different algorithms on the synthetic hyperspectral data.

| model | $L_2 - L_1$ | $L_2 - SL_0$ | $L_1 - L_1$ | $L_1 - SL_0$ |
|---|---|---|---|---|
| $rmse$ | 0.0751 | 0.0329 | 0.0255 | 0.0222 |

$x^0 = arg\min_x \|y - Ax\|_2$. In the $L_1 - SL_0$ model and $L_1 - L_1$ model, we set the initial solution $x^0 = arg\min_x \|y - Ax\|_1$.

The results obtained on the synthetic data are shown in Tab. 1. We can find that the $L_1 - *$ ($*$ represents $SL_0$ or $L_1$) model behaves better than the respective $L_2 - *$ model. This result indicates that minimization of the $L_1$ norm of the error term can improve the unmixing results compared with the minimization of the $L_2$ norm of the error term. Besides, it can be also observed that the $L_i - SL_0$(i=1,2) model behaves better than the respective $L_i - L_1$(i=1,2) model. This observation demonstrates that the smoothed $L_0$ norm regularization can better approximate the sparsest solution than the $L_1$ regularization. Among all the models, the $L_1 - SL_0$ model gets the best performance. Fig. 4 displays the true abundance maps as well as the abundance maps estimated by the four models.

*4.2. Experiments With Real Data*

In our real data experiment, we use a subimage of the AVIRIS Cuprite dataset [1] with $70 \times 70$ pixels. The scene consists of 224 spectral bands between $0.4\mu m$ and $2.5\mu m$, with spectral resolution of $10nm$. Prior to the analysis, bands $1-2$, $105-115$, $150-170$ and $223-224$ were removed due to water absorption and low SNR in those bands, leaving a total of 188 spectral bands. The Cuprite site is well understood mineralogically, and has several exposed minerals of interest, all included in the USGS library. The spectral library used here is the same with that in the synthetic experiment (including 498 different spectral signatures) with the corresponding bands removed. The mineral map which was produced by a Tricorder 3.3 software product [2] is shown in Fig. 5. Note that the Tricorder map was produced in 1995, while the publicly available AVIRIS Cuprite data were collected in 1997. Thus, we only adopt the mineral maps as a reference to make a qualitative analysis of the performances of the different sparse unmixing methods.

As the $L_2 - SL_0$ model, the $L_1 - SL_0$ model and the $L_1 - L_1$ model behave much better than the $L_2 - L_1$ model in the synthetic data experiment, we only display the results obtained by these three best models. The $\lambda$'s in the $L_2 - SL_0$ model, the $L_1 - SL_0$ model and the $L_1 - L_1$ model are set to be 0.2, 0.2 and 1, respectively. Fig. 6 shows a qualitative comparison between the fractional abundance maps of 3 highly materials in the considered scene estimated by the three models. The distribution maps of these materials produced by Tricorder software are also displayed. Note that the Tricorder maps and abundance maps estimated by the sparse unmixing algorithms are indeed different. Tricorder maps consider each pixel in the hyperspectral imagery pure and classify it as member of a specific class correlated to the representative mineral in the pixel. By contrast, as unmixing can be regarded as a classification process at the subpixel level, the abundances for a mixed pixel depend on the degree of presence of the mineral in the pixel. This distinction can account for why the Tricorder maps and the abundance maps can not be exactly the same. However, we can observe

[1]http://aviris.jpl.nasa.gov/html/aviris.freedata.html
[2]http://speclab.cr.usgs.gov/PAPERS/tetracorder/

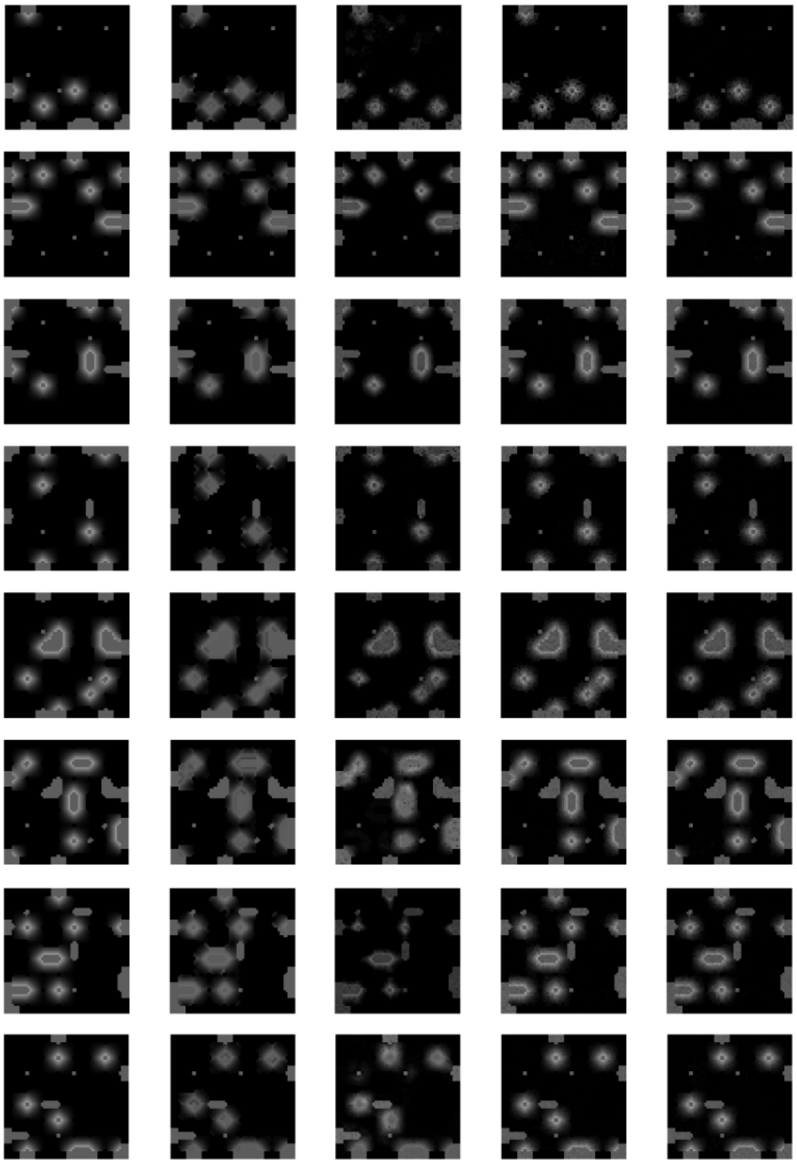

Figure 4: True abundance maps and abundance maps obtained by the four models on the synthetic data. From the left column to the right column are true abundance maps and the results obtained by the $L_2 - SL_0$ model, the $L_2 - L_1$ model, the $L_1 - SL_0$ model and the $L_1 - L_1$ model, respectively. From the top row to the bottom row are the abundance maps corresponding to Rhodochrosite HS67, Axinite HS342.3B, Chrysocolla HS297.3B, Niter GDS43 (K-Saltpeter), Anthophyllite HS286.3B, Neodymium_Oxide GDS34, Monazite HS255.3B and Samarium_Oxide GDS36, respectively.

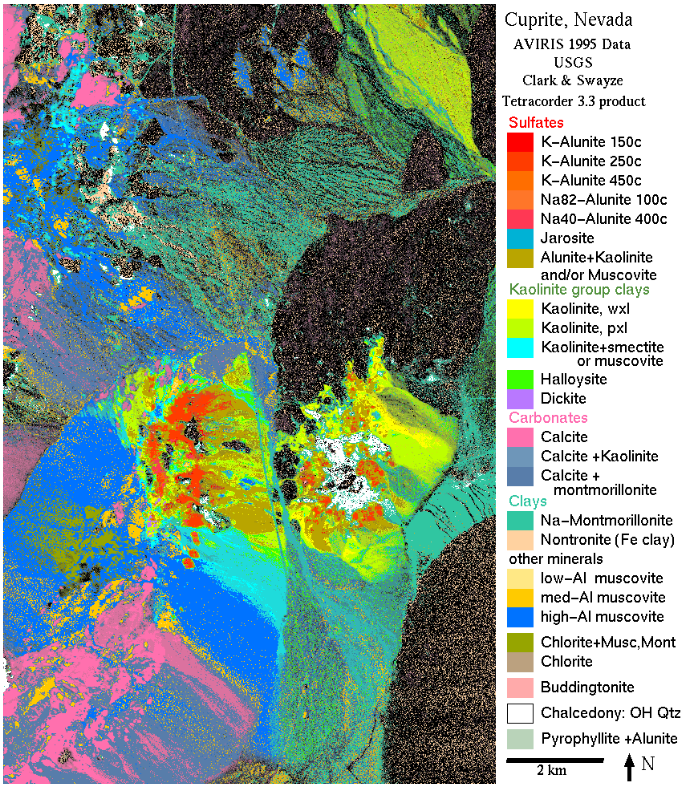


Figure 5: USGS map showing the distribution of different minerals in the Cuprite mining district in Nevada.

that the highest abundances estimated by the sparse unmixing algorithms generally correspond with those pixels classified as members of the respective class of materials. From a qualitative point of view, we can conclude that the $L_2 - SL_0$ model, the $L_1 - SL_0$ model and the $L_1 - L_1$ model are all valid to unmix the real world hyperspectral data.

## 5. Conclusion

In this paper, to solve the sparse unmixing problem, we consider using the $L_1$ norm to replace the conventional $L_2$ norm to measure the error tolerance and propose a new smooth function to approximate the $L_0$ norm to measure the sparsity. Experimental results on both the synthetic data and real data both demonstrate the efficacy of the new models.

## References


[1] N. Keshava, J. F. Mustard, Spectral unmixing, IEEE Signal Processing Magazine 19 (1) (2002) 44–57.

[2] J. P. V. P. Pauca, R. J. Plemmons, Nonnegative matrix factorization for spectral data analysis, Linear Algebra Appl. 416 (1) (2006) 29–47.

[3] W. M. A. Ambikapathi, T. Chan, C. Chi, Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing, IEEE Transactions on Geoscience and Remote Sensing 49 (11) (2011) 4194–4209.

[4] R. P. A. Plaza, P. Martinez, J. Plaza, A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data, IEEE Transactions on Geoscience and Remote Sensing 42 (3) (2004) 650–663.

[5] M. Winter, Fast autonomous spectral end-member determination in hyperspectral data, in: Proc. $13^{th}$ Int. Conf. Appl. Geologic Remote Sens., Vol. 2, Vancouver, BC, Canada, 1999, pp. 337–344.

[6] J. Boardman, Automating spectral unmixing of aviris data using convex geometry concepts, in: Proc. Summ. 4th Annu. JPL Airborne Geosci. Workshop, Vol. 1, 1993, pp. 11–14.

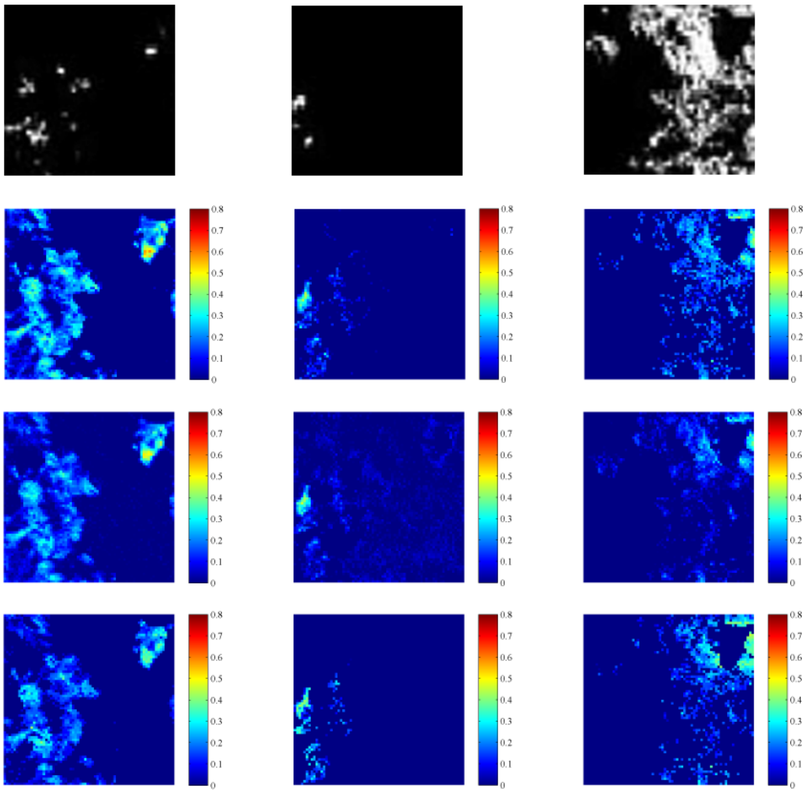

Figure 6: The fractional abundance maps estimated by $L_1 - SL_0$ model, $L_1 - L_1$ model, $L_2 - SL_0$ model and the distribution maps produced by Tricorder software for the $70 \times 70$ pixels subset of AVIRIS Cuprite scene. From top row to bottom row are the maps produced or estimated by Tricorder software, $L_1 - SL_0$ model, $L_1 - L_1$ model, $L_2 - SL_0$ model, respectively. From left column to right column are the maps corresponding to Alunite, Buddingtonite and Chalcedony, respectively.

[7] J. M. Nascimento, J. Bioucas-Dias, Vertex component analysis: A fast algorithm to unmix hyperspectral data, IEEE Trans. Geosci. Remote Sens. 43 (4) (2005) 898–910.

[8] Y. M. H. T. H. Chan, C. Y. Chi, W. K. Ma, A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing, IEEE Transactions on Geoscience and Remote Sensing 47 (11) (2009) 4418–4432.

[9] J. Li, J. Bioucas-Dias, Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data, in: IEEE Geoscience and Remote Sensing Symposium- IGARSS'08, Boston, 2008.

[10] R. L. A. E. R. D. M. Berman, H. Kiiveri, J. F. Huntington, Ice: a statistical approach to identifying endmembers in hyperspectral images, IEEE Trans. Geosci. Remote Sens. 42 (10) (2004) 2085–2095.

[11] Y. A. Abderrahim Halimi, N. Dobigeon, Nonlinear unmixing of hyperspectral images using a generalized bilinear model, IEEE Trans. Geosci. Remote Sens. 49 (11) (2011) 4153–4162.

[12] M. O. S. J. B. Adams, P. E. Johnson, Spectral mixture modeling: a new analysis of rock and soil types at the viking lander 1 site, Journal of Geophysical Research 91 (1986) 8098–8112.

[13] D. C. Heinz, C. Chang, Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery, IEEE Trans. Geosci. Remote Sens 39 (3) (2001) 529–545.

[14] J. Settle, On the relationship between spectral unmixing and subspace projection, IEEE Trans. Geosci. Remote Sens 34 (4) (1996) 1045–1046.

[15] J. T. N. Dobigeon, C. Chang, Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery, IEEE Trans. Signal Process 56 (7) (2008) 2684–2695.

[16] C. M. O. Eches, N. Dobigeon, J. Tourneret, Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery, IEEE Trans. Image Process 19 (6) (2010) 1403–1413.

[17] P. Paatero, U. Tapper, Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (2) (1994) 116–126.

[18] D. D. Lee, H. S. Seung, Learning the parts of objects by nonnegative matrix factorization, Nature 401 (6755) (1999) 788–791.

[19] L. Miao, H. Qi, Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization, IEEE Trans. Geosci. Remote Sens. 45 (3) (2007) 765–777.

[20] J. Z. Yuntao Qian, Sen Jia, A. Robles-Kelly, Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization, IEEE Trans. Geosci. Remote Sens. 49 (11) (2011) 4282–4297.

[21] Y. Y. P. Y. Xiaoqiang Lu, Hao Wu, X. Li, Manifold regularized sparse nmf for hyperspectral unmixing, IEEE Trans. Geosci. Remote Sens. (2012) 1–12.

[22] J. M. B.-D. Marian-Daniel Iordache, A. Plaza, Sparse unmixing of hyperspectral data, IEEE Trans. Geosci. Remote Sens. 49 (6) (2011) 2014–2039.

[23] J. M. B.-D. Marian-Daniel Iordache, A. Plaza, Total variation spatial regularization for sparse hyperspectral unmixing, IEEE Transactions on Geoscience and Remote Sensing 50 (11) (2012) 4484–4502.

[24] T.-Z. H. M. K. N. Xi-Le Zhao, Fan Wang, R. Plemmons, Deblurring and sparse unmixing for hyperspectral images, IEEE Trans. Geosci. Remote Sens.Article in press.

[25] J. Bioucas-Dias, A. Plaza, Collaborative sparse unmixing of hyperspectral data, in: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, Munich, 2012, pp. 7488–7491.

[26] S. Portnoy, R. Koenker, The gaussian hare and the laplacian tortoise: Computability of squared-error versus absolute-error estimators, Statistical Sci. 12 (4) (1997) 279–300.

[27] A. Giloni, M. Padberg, Alternative methods of linear regression, Math Comput Modelling 35 (2002) 361–374.

[28] H. Z. Z.f. Cui, W. Lu, An improved smoothed $l_0$-norm algorithm based on multiparameter approximation function, in: IEEE Innternational Conference on Communicartion Technology (ICCT), 2010, pp. 942–945.

[29] R. T. T. Hastie, J. Friedman, The Elements of Statistical Learnin, springer, 2008.

[30] A. C. G. J. A. Tropp, M. J. Strauss, Algorithms for simultaneous sparse approximation. part i: Greedy pursuit, Signal Processing 86 (3) (2006) 572–588.

[31] R. R. Y. C. Pati, P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in: Proc. 27th Asilomar Conference, Los Alamitos, CA, USA, 2003.

[32] J. Chen, X. Huo, Theoretical results on sparse representations of multiple-measurement vectors, IEEE Transactions on Signal Processing 54 (12) (2006) 4634–4643.

[33] J. Chen, X. Huo, Sparse representations for multiple measurement vectors (mmv) in an overcomplete dictionary, in: ICASSP 2005, 2005, pp. 257–260.

[34] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Transactions on information theory 55 (5) (2009) 2230–2249.

[35] D. D. S. Chen, M. Saunders, Atomic decomposition by basis pursuit, SIAM review 43 (1) (2001) 129–159.

[36] M. B.-Z. Hosein Mohimani, C. Jutten, A fast approach for overcomplete sparse decomposition based on smoothed l0 norm, IEEE Transactions on Signal Processing 57 (1) (2009) 289–301.

[37] H. Z. Z. B. Xu, Y. Wang, $l_{1/2}$ regularization, Sci. China Inf. Sci. 53 (6) (2010) 1159–1169.

[38] H. Zou, R. Z. Li, One-step sparse estimates in nonconcave penalized likelihood models, Ann. Statistic 36 (2008) 1509–1533.

[39] M. J.-J. I. F. F. Stamey TA, Kabalin JN, Y. N. Redwine EA, Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. ii. radical prostatectomy treated patients, J Urol. 141 (5) (1989) 1076–1083.

[40] A. G. T. V. K. R. N. Clark, G. A. Swayze, W. M. Calvin, The u.s. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns, Open file rep., U.S. Geol. Surv., Denver (1993).