Collaborative Sparse Hyperspectral Unmixing Using l_0 norm

Zhenwei Shi*, Tianyang Shi, Min Zhou, Xia Xu

***Abstract*—Sparse unmixing has been applied on hyperspectral imagery popularly in recent years. It assumes that every observed signature is a linear combination of just a few spectra (endmembers) from a known spectral library. However, solving the sparse unmixing problem directly (using l_0 norm to control the sparsity of solution at a low level) is NP-hard. Most related works focus on convex relaxation methods, but the sparsity and accuracy of results cannot be well guaranteed. Under these circumstances, this paper proposes a novel algorithm termed as collaborative sparse hyperspectral unmixing using l_0 norm (CSUnL0), which aims at solving l_0 problem directly. Firstly, it introduces a row-hard-threshold function. The row-hard-threshold function makes it possible to combine l_0 norm, instead of its approximate norms, with alternating direction method of multipliers (ADMM). Compared with convex relaxation methods, the l_0 norm constraint guarantees sparser and more accurate results. Moreover, the anti-noise ability of CSUnL0 also gets improved. Secondly, CSUnL0 uses l_2 norm of each endmembers' abundances across the whole map as a collaborative constraint, which can take advantage of the hyperspectral data's subspace property. The experimental results indicate that l_0 norm contributes to acquiring a more sparser solution and helps CSUnL0 to enhance calculation accuracy.**

***Index Terms*—Hyperspectral image, Collaborative sparse unmixing, l_0 norm, Alternating direction method of multipliers (ADMM).**

I. Introduction

With the development of technology, each pixel of a hyperspectral image has hundreds of amplitude values corresponding to its wave bands. So it is realizable to draw continuous spectral curves for different materials, which is of great help in image detection and identification. However, under the limit of low space resolution of hyperspectral images, each pixel is a mixture of several spectra corresponding to distinct materials [1]. Hence, it is in need of identifying the component spectra (endmembers) and estimating their corresponding fractions (abundances) [2], [3].

In recent years, for simplicity and efficiency, the linear model has been widely applied to hyperspectral unmixing problem. It assumes that the signature of each pixel is a linear combination of distinct endmembers [4], [5]. Under this model, there are three kinds of main primitive unmixing methods respectively based on geometry [6]–[8], statistics [9], and nonnegative matrix factorization (NMF) [10]–[12]. They are all unsupervised methods and can extract endmembers directly from hyperspectral images. However, these methods could gain virtual endmembers without any physical meaning [13], or suppose there is at least one pure pixel existing in the image [6]–[8] (It is hard to guarantee this assumption in real hyperspectral images).

In order to overcome the aforementioned disadvantages and make full use of pre-existing spectral libraries, sparse unmixing model [14]–[16] was introduced and has been used in hyperspectral unmixing. It utilizes the spectral libraries as *a priori* knowledge and assumes all endmembers in a hyperspectral image can be found in a known large spectral library. Furthermore, the number of endmembers in the hyperspectral image is much smaller than the number of spectra in the library. That is to say, sparse unmixing problem can be described as an optimization problem which aims at finding the sparsest solution (equivalent to minimize l_0 norm of an abundance vector) based on the hyperspectral images and physical constraints. However, this optimization problem is NP hard [17].

For sparse unmixing, there are five major types of calculative methods [18], namely, brute force methods [19], sparse Bayesian methods [20], greedy pursuit methods [21]–[24], nonconvex optimization methods [17], [25], [26] and convex relaxation methods [27]–[29]. The brute force methods traverse all possible solutions to get the best one and cutting-plane method can reduce the number of possible solutions. The sparse Bayesian methods, based on Bayesian statistics, utilize the sparsity and non-negativity of the abundance vector as *a priori* to calculate the maximum *a posteriori* estimation of the abundance vector. Meanwhile, the sparse Bayesian methods have very high computational complexity. The greedy pursuit methods try to find a locally optimal solution in each step and finally reach a global optimum. They are simpler than Bayesian methods. However, due to the strong correlation of endmembers in the spectral library, greedy pursuit methods could encounter the problem of trapping into local optimum sometimes. The nonconvex optimization methods attempt to relax l_0 problem to a related nonconvex problem, but it brings high computational complexity. The convex relaxation methods often replace the l_0 norm with the l_1 norm or others, and these methods can get further improved by considering *priori* information [30] or the influence of spectral variation [31]. Nevertheless, the l_1 norm just aims at making the sum

The work was supported by the National Key R&D Program of China under the Grant 2017YFC1405600 and the National Natural Science Foundation of China under the Grant 61671037. *(Corresponding author: Zhenwei Shi.)*

Zhenwei Shi (Corresponding Author, e-mail: shizhenwei@buaa.edu.cn), Tianyang Shi (e-mail: shitianyang@buaa.edu.cn), Min Zhou (e-mail: minzhou_eureka@buaa.edu.cn) and Xia Xu (e-mail: xuxia@buaa.edu.cn) are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China.

of fractional abundances' absolute values as small as possible, so it cannot guarantee sufficient sparseness of the resulting abundance vector sometimes [32], [33]. In addition to the above classical methods, some multi-objective based methods are also used for sparse unmixing [34].

In this paper, we propose a novel collaborative sparse hyperspectral unmixing algorithm using l_0 norm (CSUnL0) to solve the above problems. This algorithm is inspired by the collaborative sparse unmixing by variable splitting and augmented Lagrangian algorithm (CLSUnSAL) [28], which belongs to convex relaxation methods. Different from CLSUnSAL, CSUnL0 aims to solve the sparse unmixing problem directly without replacing l_0 norm by l_1 norm. In summary, there are two contributions of this paper:

1. For solving l_0 problem, CSUnL0 uses variable the splitting method and introduces a row-hard-threshold function, which makes it possible to apply alternating direction method of multipliers (ADMM [35], [36]) into situations where l_0 norm exists. Especially, we present a relevant theorem as a support for the convergence analysis of CSUnL0 algorithm.

2. Based on the above proofs, before l_0 norm calculation, CSUnL0 uses l_2 norm as a metric of the abundance vector to realize collaborative sparse unmixing, which means it directly limits the sparsity in a reasonable range by integrating global information. Thus, abundance vectors acquired by CSUnL0 are sparser. That is to say, under the novel sparse model, the anti-noise ability of this algorithm is stronger and the result is more accurate.

Therefore, under some conditions, the original NP-hard problem can be transformed into a simpler problem which aims to obtain an acceptable sub-optimal solution, and this could be solved in polynomial-time.

The rest of this paper is organized as follows. Section II mainly reviews the sparse unmixing of the hyperspectral data model. In Section III, we introduce our novel algorithm in details and present related proofs. Then, in Section IV, we present our experimental results and make some discussion. Finally, Section V is our conclusion.

II. Sparse Unmixing of Hyperspectral Data model

ACCORDING to the linear unmixing model, each pixel of a hyperspectral image can be decomposed into a linear combination of endmembers, as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{L\times 1}$ is an observed spectral vector of a pixel in the hyperspectral image with L bands, $\mathbf{A} \in \mathbb{R}^{L\times m}$ is a spectral library including m spectral vectors, $\mathbf{x} \in \mathbb{R}^{m\times 1}$ is an abundance vector including m elements, each element is mutual independence and corresponds to a spectral vector of the spectral library, $\mathbf{n} \in \mathbb{R}^{L\times 1}$ is observation noise.

Usually, the number of materials (endmembers) in a hyperspectral image is far smaller than the number of spectral vectors in the spectral library. The abundance vector $\mathbf{x}$ should have a few non-zero elements, which means the vector $\mathbf{x}$ is sparse. Furthermore, the model in Eq.(1) is supposed to meet the following two physical constraints:

$$x_i \geq 0,\ i = 1, 2, ..., m \tag{2}$$

$$\sum_{i=1}^{m} x_i = 1 \tag{3}$$

where x_i is the ith element of the vector $\mathbf{x}$. The above constraints correspond to abundance non-negativity constraint and sum-to-one constraint respectively [37]. However, in the real hyperspectral images, spectral vectors is variable. So the non-negative proportion coefficients ω is introduced to revise the sum-to-one constraint. Then the generalized sum-to-one constraint is shown as follow [38]:

$$\sum_{i=1}^{m} \omega_i x_i = 1 \tag{4}$$

Moreover, the sparsity of the vector $\mathbf{x}$ indicates that each pixel in the hyperspectral image can be expressed in a few spectral vectors in the spectral library. Therefore, the sparse unmixing optimization problem can be described as follows:

$$\begin{aligned} &\min_{\mathbf{x}} \ ||\mathbf{x}||_0 \\ &\text{subject to}: \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2 \leq \varepsilon,\ \mathbf{x} \geq 0 \end{aligned} \tag{5}$$

where $||\mathbf{x}||_0$ is l_0 norm representing the number of non-zero elements in the vector $\mathbf{x}$, and ε is an error margin. Because l_0 norm is discrete and nonconvex, the optimization problem of l_0 norm is NP-hard [17].

A popular method is to replace the l_0 norm with l_1 norm which is a relative ideal convex approximation of l_0 norm [39]. Then the new sparse unmixing optimization problem can be written as:

$$\begin{aligned} &\min_{\mathbf{x}} \ ||\mathbf{x}||_1 \\ &\text{subject to}: \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2 \leq \varepsilon,\ \mathbf{x} \geq 0 \end{aligned} \tag{6}$$

Furthermore, the application range of the above formulas can be extended from a pixel to the whole hyperspectral image. Then a collaborative sparse unmixing model can be proposed [28], which takes advantage of the hyperspectral data's subspace property. Assuming that $\mathbf{Y} \in \mathbb{R}^{L\times K}$ is a hyperspectral image matrix including L bands and K pixels, each column vector is a spectral vector; $\mathbf{A} \in \mathbb{R}^{L\times m}$ is a spectral library including m spectral vectors; $\mathbf{X} \in \mathbb{R}^{m\times K}$ is an abundance matrix, whose elements correspond to abundance of m spectra in K pixels. Meanwhile, the matrix $\mathbf{X}$ should be row sparse.

Collaborative linear unmixing model can be written as:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{7}$$

where $\mathbf{N} \in \mathbb{R}^{L\times K}$ is an observation noise matrix.

As mentioned above, the following optimization problem can be solved to estimate the abundance of the endmembers:

$$\begin{aligned} &\min_{\mathbf{X}} \ ||\mathbf{X}||_{r-0} \\ &\text{subject to}: \ ||\mathbf{Y} - \mathbf{A}\mathbf{X}||_F \leq \varepsilon,\ \mathbf{X} \geq 0 \end{aligned} \tag{8}$$

where $||\cdot||_F$ is the Frobenius norm of matrix, $||\mathbf{X}||_{r-0}$ (row-zero norm) means the number of non-zero rows in $\mathbf{X}$ [40].

Similarly, above optimization problem is also NP-hard. Therefore, $||\mathbf{X}||_{r-0}$ is usually relaxed to $||\mathbf{X}||_{2,1}$ ($||\mathbf{X}||_{2,1} = \sum_{i=1}^{m} ||\mathbf{x}_i||_2$, where $\mathbf{x}_i$ is ith row in $\mathbf{X}$) which is $l_{2,1}$ mixed norm [41]. As a consequence, Eq.(8) can be rewritten as follows:

$$\begin{aligned} &\min_{\mathbf{X}} \ ||\mathbf{X}||_{2,1} \\ &\text{subject to : } ||\mathbf{Y} - \mathbf{AX}||_F \leq \varepsilon, \ \mathbf{X} \geq 0 \end{aligned} \tag{9}$$

By introducing a suitable Lagrange multiplier, the collaborative linear sparse unmixing model can be written as:

$$\begin{aligned} &\min_{\mathbf{X}} \ \frac{1}{2}||\mathbf{Y} - \mathbf{AX}||_F^2 + \lambda||\mathbf{X}||_{2,1} \\ &\text{subject to : } \mathbf{X} \geq 0 \end{aligned} \tag{10}$$

The $\lambda \geq 0$ is the regularization parameter and here is used for adjusting the sparsity of solution. Then the above problem can be solved by CLSUnSAL [28] which is based on ADMM.

However, the tests have shown that the solution of optimization problem under the $l_{2,1}$ norm constraint cannot keep sparse all the time, that is to say, it could produce more virtual endmembers which are inactive in the hyperspectral image [32], [33]. Then we expect a kind of constraint which can reduce virtual endmembers. Inspired by iterative hard-thresholding algorithm [42], we look back to l_0 norm to get a more accurate solution.

III. Collaborative Sparse Hyperspectral Unmixing Using l_0 norm

In this section, firstly, we introduce a new sparse unmixing model (CSUnL0) which adopts an $l_{2,0}$ norm constraint rather than the $l_{2,1}$ norm constraint. Secondly, we use ADMM to solve the corresponding optimization problem.

A. CSUnL0 model description

The motivations of the new sparse unmixing model are explained as follows: Firstly, the l_1 norm is an approximation of the l_0 norm and the corresponding algorithm is difficult to achieve high accuracy. Secondly, we found that the closed form solution of the row-zero norm problem can be achieved by splitting variables and using a row-hard-threshold function.

For illustration purpose, the row-hard-threshold function is defined as:

$$(\mathcal{RH}_t(\mathbf{X}))_{i,j} = \begin{cases} x_{i,j} & \sum_j x_{i,j}^2 > t \\ 0 & \sum_j x_{i,j}^2 \leq t \end{cases} = 1(\sum_j x_{i,j}^2 > t) \cdot x_{i,j} \tag{11}$$

where t is a threshold and $x_{i,j}$ represents the jth element of the ith row in the matrix $\mathbf{X}$. Moreover, $1(\cdot)$ is one if the proposition in the bracket is true, and zero otherwise.

Lemma 1. *For the optimization problem:*

$$\min_{\mathbf{\Phi}} \ ||\mathbf{\Psi} - \mathbf{\Phi}||_F^2 + \lambda||\mathbf{\Phi}||_{r-0} \tag{12}$$

Its closed-form solution is:

$$\mathbf{\Phi} = \mathcal{RH}_\lambda(\mathbf{\Psi}) \tag{13}$$

Meanwhile, l_2 norm is used as a metric which decides whether to retain a row of $\mathbf{\Psi}$.

Proof. The proof of Lemma 1 is given in APPENDIX A. □

Based on the Lemma 1, we can use l_2 norm as a metric of non-zero rows, then the new sparse unmixing model from the Eq.(8) can be written as:

$$\min_{\mathbf{X}} \ \frac{1}{2}||\mathbf{Y} - \mathbf{AX}||_F^2 + \lambda||\mathbf{X}||_{2,0} + l_{R+}(\mathbf{X}) \tag{14}$$

where we define $||\mathbf{X}||_{2,0} = \sum_{i=1}^{m} 1(||\mathbf{x}_i||_2 > 0)$ and the matrix $\mathbf{X}$ contains m endmembers and K pixels. Besides, $l_{R+}(\mathbf{X})$ is zero if $\mathbf{X} \geq 0$, and $+\infty$ otherwise.

Then we introduce three variables $\mathbf{V}_1$,$\mathbf{V}_2$ and $\mathbf{V}_3$ to split variables, the Eq.(14) is equivalent to:

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{V_1},\mathbf{V_2},\mathbf{V_3}} \frac{1}{2}||\mathbf{Y} - \mathbf{V}_1||_F^2 + \lambda||\mathbf{V}_2||_{2,0} + l_{R+}(\mathbf{V}_3) \\ &\text{subject to } : \ \mathbf{V}_1 = \mathbf{AX} \\ &\qquad\qquad\quad \mathbf{V}_2 = \mathbf{X} \\ &\qquad\qquad\quad \mathbf{V}_3 = \mathbf{X} \end{aligned} \tag{15}$$

The augmented Lagrange function of the Eq.(15) can be written as:

$$\begin{aligned} &L(\mathbf{X}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3) \\ =&\frac{1}{2}||\mathbf{Y} - \mathbf{V}_1||_F^2 + \lambda||\mathbf{V}_2||_{2,0} + l_{R+}(\mathbf{V}_3) + \\ &\frac{\mu}{2}||\mathbf{AX} - \mathbf{V}_1 - \mathbf{D}_1||_F^2 + \frac{\mu}{2}||\mathbf{X} - \mathbf{V}_2 - \mathbf{D}_2||_F^2 + \\ &\frac{\mu}{2}||\mathbf{X} - \mathbf{V}_3 - \mathbf{D}_3||_F^2 \end{aligned} \tag{16}$$

where $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ are Lagrangian multipliers, μ is a penalty parameter [43].

The term $\lambda||\mathbf{V}_2||_{2,0}$ in Eq.(15) is the only difference between our model and CLSUnSAL model, so we can only discuss the $\mathbf{V}_2$ part. Ignore variables independent of $\mathbf{V}_2$, the optimization problem in Eq.(16) becomes:

$$\min_{\mathbf{V}_2} \frac{\mu}{2}||\mathbf{X} - \mathbf{V}_2 - \mathbf{D}_2||_F^2 + \lambda||\mathbf{V}_2||_{2,0} \tag{17}$$

and this has the same form with the problem in Lemma 1. That is to say, the new sparse unmixing model can be solved by using ADMM which has been verified as a very effective algorithm to solve sparse unmixing problems [27], [28].

Physically, an active endmember corresponds to a material which exists in the hyperspectral image and l_2 norm of its abundance vector is relatively large. Meanwhile, because of noise's randomness and weak intensity, the l_2 norm of an inactive endmember's abundance vector caused by noise should be relatively small. Therefore, the row-hard-threshold is used to distinguish between the above two cases and exclude inactive endmembers during iteration.

B. CSUnL0 algorithm

In this part, we will show in detail how ADMM is used to solve Eq.(16).

The major steps of ADMM are: update $\mathbf{X}$, $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ as optimization variables and then update Lagrange multipliers $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ one by one. The algorithm will be detailed as follows:

When updating $\mathbf{X}$, variables independent of $\mathbf{X}$ need not be taken into account. The optimization problem becomes:

$$\min_{\mathbf{X}}(\frac{\mu}{2}||\mathbf{AX}-\mathbf{V}_1^{(k)}-\mathbf{D}_1^{(k)}||_F^2+ \frac{\mu}{2}||\mathbf{X}-\mathbf{V}_2^{(k)}-\mathbf{D}_2^{(k)}||_F^2+\frac{\mu}{2}||\mathbf{X}-\mathbf{V}_3^{(k)}-\mathbf{D}_3^{(k)}||_F^2) \tag{18}$$

The above formula has a closed-form solution:

$$\begin{aligned}\mathbf{X}^{(k+1)} =&(\mathbf{A}^T\mathbf{A}+2\mathbf{I})^{-1}[\mathbf{A}^T(\mathbf{V}_1^{(k)}+\mathbf{D}_1^{(k)})+ \\ &\mathbf{V}_2^{(k)}+\mathbf{D}_2^{(k)}+\mathbf{V}_3^{(k)}+\mathbf{D}_3^{(k)}]\end{aligned} \tag{19}$$

where $\mathbf{I}$ is a unit matrix with appropriate size.

Similarly, just concentrate on one variable $\mathbf{V}_1$ and consider other variables independent of the $\mathbf{V}_1$ in Eq.(16) as constants. Then the following optimization problem can be solved:

$$\min_{\mathbf{V}_1}\frac{1}{2}||\mathbf{Y}-\mathbf{V}_1||_F^2+\frac{\mu}{2}||\mathbf{AX}^{(k+1)}-\mathbf{V}_1-\mathbf{D}_1^{(k)}||_F^2 \tag{20}$$

The above formula has a closed-form solution:

$$\mathbf{V}_1^{(k+1)}=\frac{1}{1+\mu}[\mathbf{Y}+\mu(\mathbf{AX}^{(k+1)}-\mathbf{D}_1^{(k)})] \tag{21}$$

Then it is the turn of $\mathbf{V}_2$ to become the only considered variable in Eq.(16), and define manual adjustable parameter $a=\frac{2\lambda}{\mu}$ for controlling directly, because μ is variable and incremental during optimization procedure. The optimization problem becomes:

$$\min_{\mathbf{V}_2}\frac{\mu}{2}(||\mathbf{X}^{(k+1)}-\mathbf{V}_2-\mathbf{D}_2^{(k)}||_F^2+a||\mathbf{V}_2||_{2,0}) \tag{22}$$

Based on the Lemma 1, the above formula can be solved by row-hard-threshold algorithm: $\mathbf{V}_2^{(k+1)}=\mathcal{RH}_a(\mathbf{X}^{(k+1)}-\mathbf{D}_2^{(k)})$, as shown in the Algorithm 1:

Algorithm 1 Pseudo-code for row-hard-threshold algorithm

Input: $\mathbf{X}\in\mathbb{R}^{m\times K}, a\geq 0$ (where a is a threshold value)
for $i=1...m$
  if $||\mathbf{x}_i||_2^2\leq a$ ($\mathbf{x}_i$ is the ith row of $\mathbf{X}$)
    $\mathbf{x}_i=\mathbf{0}$;
  end if
end for
Output: $\mathbf{X}$

After updating $\mathbf{V}_2$, we do the same to update $\mathbf{V}_3$ by solving the optimization problem below:

$$\min_{\mathbf{V}_3}\frac{\mu}{2}||\mathbf{X}^{(k+1)}-\mathbf{V}_3-\mathbf{D}_3^{(k)}||_F^2+l_{R+}(\mathbf{V}_3) \tag{23}$$

The above formula is used to project $\mathbf{X}^{(k+1)}-\mathbf{D}_3^{(k)}$ into non-negative quadrant:

$$\mathbf{V}_3^{(k+1)}=\max(\mathbf{X}^{(k+1)}-\mathbf{D}_3^{(k)},0) \tag{24}$$

The updating procedure of Lagrangian multipliers $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ is executed after updating $\mathbf{X}$, $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$. The details are described in the following.

The complete CSUnL0 algorithm is summarized in Algorithm 2:

Algorithm 2 Pseudo-code for CSUnL0

1: **Initialization**:
2: set $k=0$, choose $a>0, \mu>0, \epsilon>0$, $\mathbf{V}_1^0,\mathbf{V}_2^0,\mathbf{V}_3^0,\mathbf{D}_1^0,\mathbf{D}_2^0,\mathbf{D}_3^0$ ($\mathbf{Y}\in\mathbb{R}^{L\times K}$,$\mathbf{A}\in\mathbb{R}^{L\times m}$,$\epsilon$ is the normalized error tolerance)
3: **Repeat**:
4: $\mathbf{X}^{(k+1)}=(\mathbf{A}^T\mathbf{A}+2\mathbf{I})^{-1}[\mathbf{A}^T(\mathbf{V}_1^{(k)}+\mathbf{D}_1^{(k)})+\mathbf{V}_2^{(k)}+\mathbf{D}_2^{(k)}+\mathbf{V}_3^{(k)}+\mathbf{D}_3^{(k)}]$
5: $\mathbf{V}_1^{(k+1)}=\frac{1}{1+\mu}[\mathbf{Y}+\mu(\mathbf{AX}^{(k+1)}-\mathbf{D}_1^{(k)})]$
6: $\mathbf{V}_2^{(k+1)}=\mathcal{RH}_a(\mathbf{X}^{(k+1)}-\mathbf{D}_2^{(k)})$
7: $\mathbf{V}_3^{(k+1)}=\max(\mathbf{X}^{(k+1)}-\mathbf{D}_3^{(k)},0)$
8: $\mathbf{D}_1^{(k+1)}=\mathbf{D}_1^{(k)}-\mathbf{AX}^{(k+1)}+\mathbf{V}_1^{(k+1)}$
9: $\mathbf{D}_2^{(k+1)}=\mathbf{D}_2^{(k)}-\mathbf{X}^{(k+1)}+\mathbf{V}_2^{(k+1)}$
10: $\mathbf{D}_3^{(k+1)}=\mathbf{D}_3^{(k)}-\mathbf{X}^{(k+1)}+\mathbf{V}_3^{(k+1)}$
11: $k=k+1$
12: **Until**: $||\mathbf{V}_1^k-\mathbf{AX}^k||_F+||\mathbf{V}_2^k-\mathbf{X}^k||_F+||\mathbf{V}_3^k-\mathbf{X}^k||_F<\sqrt{(2m+L)\times K}\epsilon$ or the maximum iteration is reached.

From Algorithm 2, it is shown that the complexity of CSUnL0 depends on updating $\mathbf{X}$ in the fourth line, which has the same form with CLSUnSAL. So CSUnL0 and CLSUnSAL have the same complexity [28]. Furthermore, assume that there are L bands, K pixels, and m endmembers in the hyperspectral image. Because the term $(\mathbf{A}^T\mathbf{A}+2\mathbf{I})^{-1}$ is constant, the complexity of updating $\mathbf{X}$ depends on the other matrix multiplications which is $\max(\mathcal{O}(mLK),\mathcal{O}(m^2K))$.

Due to the non-convexity of CSUnL0, we will discuss the convergence condition of our algorithm in Note 1 and Theorem 1.

Note 1. *Assuming that the abundance matrix $\mathbf{X}$ in Algorithm 2 is always non-negative, which means $\mathbf{X}\geq 0$ is true. Then the original algorithm can be written as follows:*

$$\begin{aligned}&\min_{\mathbf{X},\mathbf{V_1},\mathbf{V_2},\mathbf{V_3}}\frac{1}{2}||\mathbf{Y}-\mathbf{V}_1||_F^2+\lambda||\mathbf{V}_2||_{2,0} \\ &\textit{subject to}\ :\ \mathbf{V}_1=\mathbf{AX} \\ &\qquad\qquad\ \mathbf{V}_2=\mathbf{X}\end{aligned} \tag{25}$$

The whole augmented Lagrange function of the Eq.(25) is:

$$\begin{aligned}L(\mathbf{X},\mathbf{V}_1,\mathbf{V}_2,\mathbf{D}_1,\mathbf{D}_2)=&\frac{1}{2}||\mathbf{Y}-\mathbf{V}_1||_F^2+\lambda||\mathbf{V}_2||_{2,0}+ \\ &\frac{\mu}{2}||\mathbf{AX}-\mathbf{V}_1-\mathbf{D}_1||_F^2+\frac{\mu}{2}||\mathbf{X}-\mathbf{V}_2-\mathbf{D}_2||_F^2- \\ &\frac{\mu}{2}||\mathbf{D}_1||_F^2-\frac{\mu}{2}||\mathbf{D}_2||_F^2\end{aligned} \tag{26}$$

Furthermore, the corresponding iterative procedure is:

$$\begin{aligned}\mathbf{X}^{(k+1)}&=(\mathbf{A}^T\mathbf{A}+\mathbf{I})^{-1}[\mathbf{A}^T(\mathbf{V}_1^{(k)}+\mathbf{D}_1^{(k)})+\mathbf{V}_2^{(k)}+\mathbf{D}_2^{(k)}] \\ \mathbf{V}_1^{(k+1)}&=\frac{1}{1+\mu}[\mathbf{Y}+\mu(\mathbf{AX}^{(k+1)}-\mathbf{D}_1^{(k)})] \\ \mathbf{V}_2^{(k+1)}&=\mathcal{RH}_a(\mathbf{X}^{(k+1)}-\mathbf{D}_2^{(k)}) \\ \mathbf{D}_1^{(k+1)}&=\mathbf{D}_1^{(k)}-\mathbf{AX}^{(k+1)}+\mathbf{V}_1^{(k+1)} \\ \mathbf{D}_2^{(k+1)}&=\mathbf{D}_2^{(k)}-\mathbf{X}^{(k+1)}+\mathbf{V}_2^{(k+1)}\end{aligned} \tag{27}$$

Besides, the stopping criterion is $||\mathbf{V}_1^k - \mathbf{A}\mathbf{X}^k||_F + ||\mathbf{V}_2^k - \mathbf{X}^k||_F < \sqrt{(m+L)\times K}\epsilon$, where ϵ is a normalized error tolerance.

Theorem 1. *Before reaching the stopping criterion, $L(\mathbf{X}^k, \mathbf{V}_1^k, \mathbf{V}_2^k, \mathbf{D}_1^k, \mathbf{D}_2^k)$ in Note 1 is decreasing with a lower bound when the following conditions hold.*

1. $\lambda \leq \frac{\mu(m+L)K\epsilon^2}{m(\frac{4}{\mu}+2\sqrt{2})^2}$
2. $\mu \geq 2$

Proof. The proof of Theorem 1 is given in APPENDIX B. □

Based on the Theorem 1, it provides a weaker convergence condition. However, this theorem cannot promise the result is globally optimal, because duality feasible condition is not always true when the objective function is non-convex especially. That is to say, as long as the abundance matrix is non-negative, an acceptable result would be found in polynomial time by solving the l_0 problem in our algorithm, and this result is often good enough. Moreover, the least squares solution is used to initialize the CSUnL0 algorithm, which will be helpful to find a better sub-optimal result.

IV. Experiment Results and Discussion

HERE we employ two synthetic and one real hyperspectral images to evaluate the performance of CSUnL0.

In the first synthetic data experiment, we aim at testing the accuracy and anti-noise ability of CSUnL0. This experiment is designed as follows: We generate a hyperspectral data set satisfying the Dirichlet distribution. Then CSUnL0 is applied to this synthetic data set. Meanwhile, for comparison, we also run some other representative algorithms in the same data set, such as SMP [23], RSFoBa [24], SUnSAL [27], CLSUnSAL [28] and ADSpLRU [29]. SMP and RSFoba are classical greedy pursuit methods, SUnSAL and CLSUnSAL are classical convex relaxation methods and ADSpLRU is the newest convex relaxation method. Finally, RMSE is used as a criterion to judge whether our algorithm is qualified or not. The ith active endmember's RMSE is defined below:

$$\text{RMSE}_i = \sqrt{\frac{1}{K}\sum_{j=1}^{K}(x_{i,j} - x_{i,j}^*)^2} \quad i = 1, 2, ..., r \tag{28}$$

where $\mathbf{X}$ comes from the result of each algorithm, $\mathbf{X}^*$ is the real abundance distribution, and there are only r endmembers used to construct the data set. Each algorithm's RMSE is the average value of all endmembers' RMSEs:

$$\text{RMSE} = \frac{1}{r}\sum_{i=1}^{r}\text{RMSE}_i \tag{29}$$

Moreover, a lower RMSE level indicates that the corresponding algorithm is better.

In the second synthetic hyperspectral data experiment, we choose CLSUnSAL to compare with CSUnL0. That is because the only difference between CSUnL0 and CLSUnSAL is the sparsity constraint – CSUnL0 uses $l_{2,0}$ norm constraint while CLSUnSAL uses $l_{2,1}$ norm constraint. In order to visually present the differences between these two algorithms, some endmembers are put into several boxes with certain abundance. Besides, we will discuss the existence of parameter a and its impact on our algorithm. The details have been shown in part B.

In the real hyperspectral data experiment, the well-known AVIRIS Cuprite data set[1] is used. For the real data, there is no true abundance distribution. So RMSE is inapplicable to estimate the performance of all the algorithms above. Therefore, we just qualitatively analyze the validity of CSUnL0 compared with SUnSAL and CLSUnSAL.

Our spectral library $\mathbf{A} \in \mathbb{R}^{224\times 498}$ is Chapter 1 of the U.S. Geological Survey (USGS) [44] digital spectral library (splib06a). The reflectance values of 498 materials are measured for 224 spectral bands uniformly distributed in the interval 0.4∼2.5 μm.

Especially, in the synthetic hyperspectral data experiments, all the algorithms used to compare with ours have been tuned to the best performance by using different parameter values. For example, we choose $\lambda = \{0, 10^{-5}, 10^{-4.5}, ..., 1\}$ for SUnSAL and keep the best result. In the real hyperspectral data experiment, these algorithms have been tuned as in their papers.

A. Parameter preferences

In the row-hard-threshold algorithm of CSUnL0, the most important parameter is a, relevant to the values of endmembers' abundance. It means that the endmember whose total abundance ($||\mathbf{x}||_2^2$) is lower than a will be treated as noise and removed in the iterative process. In other words, a controls the result's sparsity. Usually, there is a way to initialize the parameter a: Set a to a relatively small value, gradually increase a and run the program until the algorithm achieves a good result. In the detailed implementation, we use a_0 as the upper bound where $0 \leq a/K \leq a_0$ and the image has K pixels. After choosing a_0 and initializing a to a very small value, we double a when $||\mathbf{V}_2||_{2,0}$ stops changing. This way can guarantee that the algorithm gradually excludes inactive endmembers and improve the algorithm's robustness.

B. Synthetic Hyperspectral Data Experiment 1

In this section, we run each algorithm with different intensities of white noise in the first two groups and different numbers of endmembers in the third group.

In the first two groups, the data sets are used with different signal-to-noise ratio. Here the signal-to-noise ratio (SNR) is defined as follows:

$$\text{SNR} = 10\lg(||\mathbf{Y}||_2^2/||\mathbf{N}||_2^2) \tag{30}$$

where $\mathbf{Y}$ represents a synthetic hyperspectral image data with some endmembers, $\mathbf{N}$ is added with white noise.

In the first group, we choose five materials from $\mathbf{A}$, to generate the corresponding abundance matrix $\mathbf{H}$ which obeys the Dirichlet distribution in 30×30 pixels, and the sum of elements in $\mathbf{H}$ is 30×30. The abundances which are smaller than 0.7 are used to avoid the pure pixel including only one

[1] http://aviris.jpl.nasa.gov/html/aviris.freedata.html

endmember [8], [27]. Those materials' spectra are shown in Fig.1.

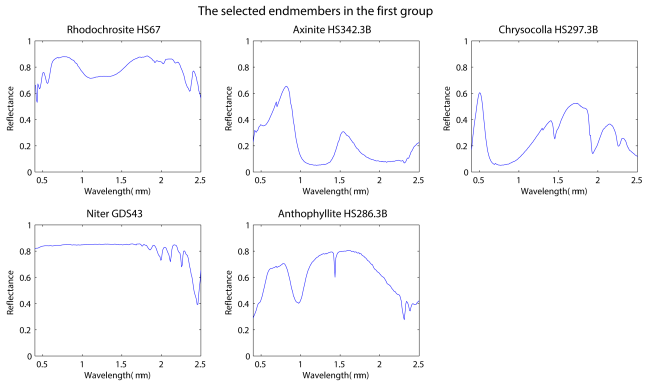

Fig. 1. Selected 5 materials' spectra. Each subgraph's title corresponds to a kind of mineral material.

So the synthetic hyperspectral image data $\mathbf{Y}$ can be generated by $\mathbf{Y} = \mathbf{AH}$. Then the $20 \sim 50$ dB white noise is added into the data set $\mathbf{Y}$. The experimental results are shown in the Fig.2 and Table I:

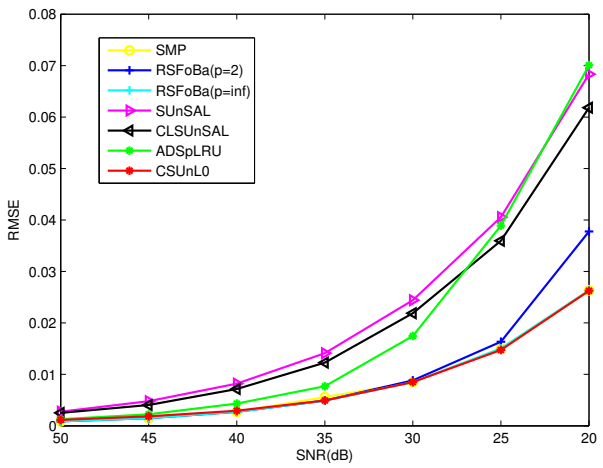

Fig. 2. Results of the first group: each algorithm's RMSE under different intensities of white noise. The results of SMP, RSFoBa (p=inf) and CSUnL0 are almost in the same value. CSUnL0 parameter setting: $a_0 = 0.04$.

It can be observed from Fig.2 that all the RMSEs increase with the intensity of white noise. The performance of CSUnL0 is much better than other convex relaxation algorithms (SUNSAL, CLSUNSAL, ADSpLRU) and very close to some greedy pursuit algorithms (SMP, RSFoBa (p=inf)). This is because CSUnL0 can exclude inactive endmembers out of the iteration process by the row-hard-threshold function, and lots of these inactive endmembers are caused by noise. Clearly, the less inactive endmembers are, the sparser and more accurate the result becomes. Therefore, it is feasible to improve the algorithm's anti-noise ability by using $l_{2,0}$ norm to replace $l_{2,1}$ norm constraint and introducing the row-hard-threshold function.

TABLE I
RMSEs OBTAINED BY EACH ALGORITHM IN THE FIRST GROUP

SNR (dB)	SMP	RSFoBa (p=2)	RSFoBa (p=inf)	SUnSAL	CLSUnSAL	ADSpLRU	CSUnL0
50	0.0009	0.0009	0.0009	0.0027	0.0025	0.0013	0.0012
45	0.0015	0.0015	0.0015	0.0048	0.0040	0.0022	0.0018
40	0.0027	0.0027	0.0027	0.0082	0.0072	0.0043	0.0029
35	0.0056	0.0049	0.0049	0.0141	0.0123	0.0077	0.0049
30	0.0085	0.0088	0.0085	0.0244	0.0219	0.0175	0.0085
25	0.0150	0.0163	0.0150	0.0406	0.0360	0.0389	0.0147
20	0.0263	0.0378	0.0263	0.0683	0.0618	0.0701	0.0262

In the second group, we choose 5 different endmembers which are shown in Fig.3. These endmembers are of Actinolite and their spectra are very similar. The rest of conditions in the second group are the same as the first group's. The experimental results are shown in Fig.4 and Table II:

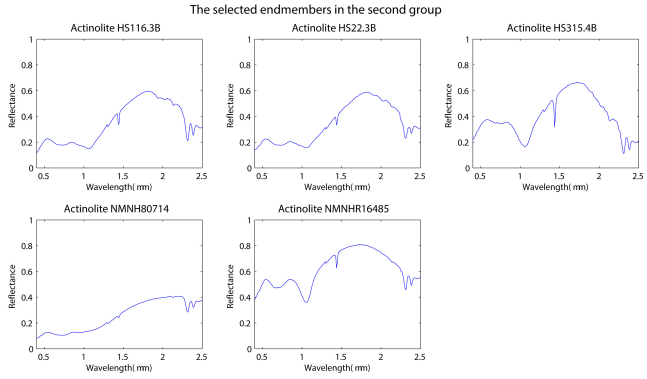

Fig. 3. Selected 5 materials' spectra with high similarity.

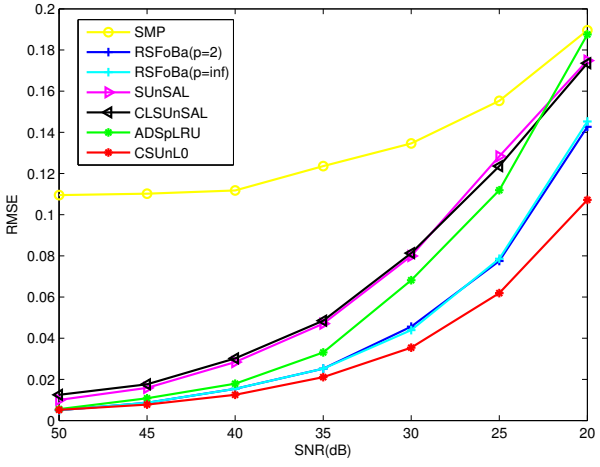

Fig. 4. Results of the second group: each algorithm's RMSE under different intensities of white noise. CSUnL0 parameter setting: $a_0 = 0.05$.

TABLE II
RMSEs OBTAINED BY EACH ALGORITHM IN THE SECOND GROUP

SNR (dB)	SMP	RSFoBa (p=2)	RSFoBa (p=inf)	SUnSAL	CLSUnSAL	ADSpLRU	CSUnL0
50	0.1095	0.0050	0.0050	0.0100	0.0125	0.0056	0.0052
45	0.1102	0.0086	0.0086	0.0159	0.0176	0.0108	0.0078
40	0.1118	0.0154	0.0154	0.0284	0.0302	0.0179	0.0125
35	0.1236	0.0252	0.0252	0.0470	0.0485	0.0331	0.0211
30	0.1346	0.0456	0.0439	0.0800	0.0813	0.0682	0.0354
25	0.1553	0.0775	0.0786	0.1284	0.1236	0.1119	0.0619
20	0.1896	0.1427	0.1453	0.1749	0.1736	0.1876	0.1072

TABLE III
RMSEs OBTAINED BY EACH ALGORITHM IN THE THIRD GROUP

n	SMP	RSFoBa (p=2)	RSFoBa (p=inf)	SUnSAL	CLSUnSAL	ADSpLRU	CSUnL0
3	0.0048	0.0048	0.0048	0.0204	0.0117	0.0058	0.0052
5	0.0089	0.0083	0.0083	0.0238	0.0215	0.0167	0.0085
7	0.0100	0.0095	0.0095	0.0217	0.0190	0.0143	0.0096
9	0.0146	0.0130	0.0122	0.0247	0.0233	0.0160	0.0099
11	0.0156	0.0185	0.0145	0.0291	0.0316	0.0240	0.0120
13	0.0203	0.0291	0.0176	0.0346	0.0355	0.0234	0.0134
15	0.0213	0.0339	0.0289	0.0378	0.0416	0.0233	0.0147

* n is the number of endmembers.

From the Fig.4, when the similarity of active endmembers increases, CSUnL0 can be better than greedy pursuit algorithms (SMP, RSFoBa). Especially, SMP cannot extract right endmembers in this case, because the selected 5 endmembers belong to the same kind of material and have very high correlation. Based on the pseudo-code of CSUnL0, it shows that CSUnL0 searches active endmembers by endmembers' abundances other than correlation, this way can avoid confusion caused by endmembers' high correlation. So CSUnL0 has a good ability in extracting endmembers.

In the third group, we choose 15 materials from spectral library $\mathbf{A}$ to create seven data sets respectively and add 30 dB white noise into the data sets. The former five endmembers are from the first group and the rest ten endmembers are Neodymium Oxide GDS34, Monazite HS255.3B, Samarium Oxide GDS36, Pinnoite NMNH123943, Meionite WS700.HLsep, Spodumene HS210.3B, Laumontite GDS5, Grossular WS484, Zoisite HS 347.3B and Wollastonite HS348.3B respectively. Results are presented in Fig.5 and Table III:

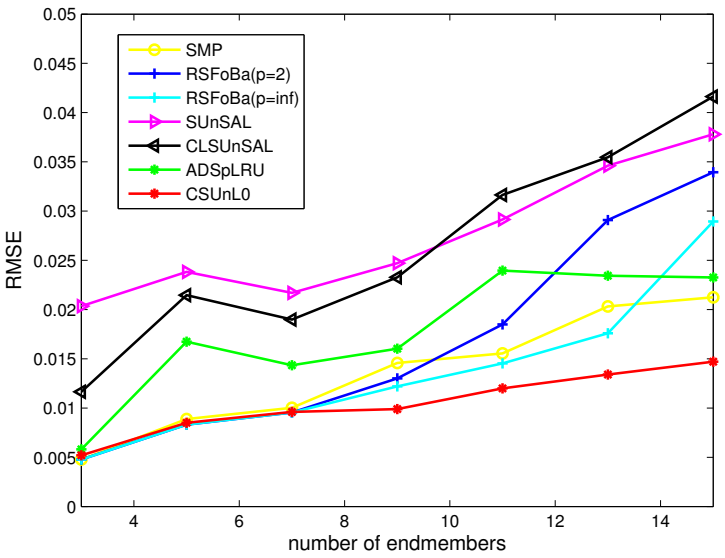

Fig. 5. Results of the third group: each algorithm's RMSE under different number of endmembers. CSUnL0 parameter setting: $a_0 = 0.02$.

From Fig.5, it can be observed that all the RMSEs increase with the number of endmembers. The performance of CSUnL0 is also better than the others, especially when the number of endmembers is large.

The above three pictures (Fig.2, Fig.4 and Fig.5) show that CSUnL0 can achieve a good anti-noise performance, a better ability in extracting endmembers and a more accurate solution. It can be observed that CSUnL0 much surpasses CSUnSAL in the above experiments. In conclusion, the introduction of $l_{2,0}$ norm constraint and the row-hard-threshold function are beneficial.

C. Synthetic Hyperspectral Data Experiment 2

This experiment is designed for visualizing algorithms' ability in extracting endmembers and anti-noise. In this experiment, we design an abundance distribution $\mathbf{H}$, which has 4×4 blocks in 45×45 pixels uniform distribution and includes five endmembers (They are Actinolite HS116.3B, Actinolite HS22.3B, Actinolite NMNH80714, Albite HS66.3B and Almandine WS477). Each block has 7×7 pixels. Each column of blocks contains four different endmembers with the same abundances, each row of blocks contains the same endmembers whose abundances are from 0.25 to 1. The fifth endmember supposed as background is set to a complementary value, which means $H_{5,j} = 1 - \sum_{i=1}^{4} H_{i,j}$ ($H_{i,j}$: the abundance of the ith endmember in jth pixel). The abundance distribution $\mathbf{H}$ is shown in Fig.6. So we can obtain the data set $\mathbf{Y} = \mathbf{AH}$. If an algorithm performs well in extracting endmembers, the position of these blocks and endmembers' abundances can be found exactly in the result. In this case, we observe how close the abundances obtained approaches the real ones. Furthermore, 30 dB and 50dB white noise are added into the data set to test the anti-noise ability of CSUnL0. The experimental results are presented in Fig.7, Fig.8 and Table IV :

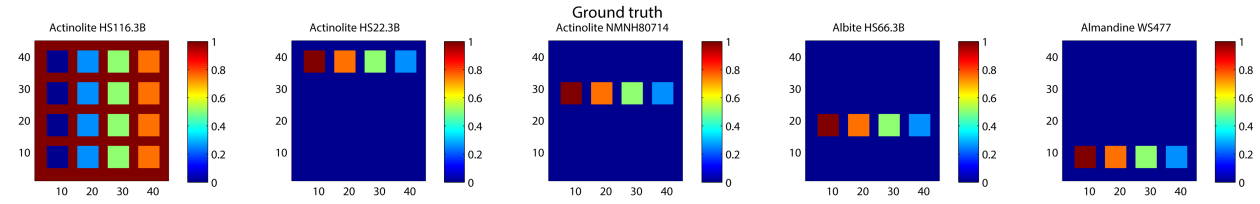

Fig. 6. Ground truth: The abundances of selected materials in the USGS spectral library.

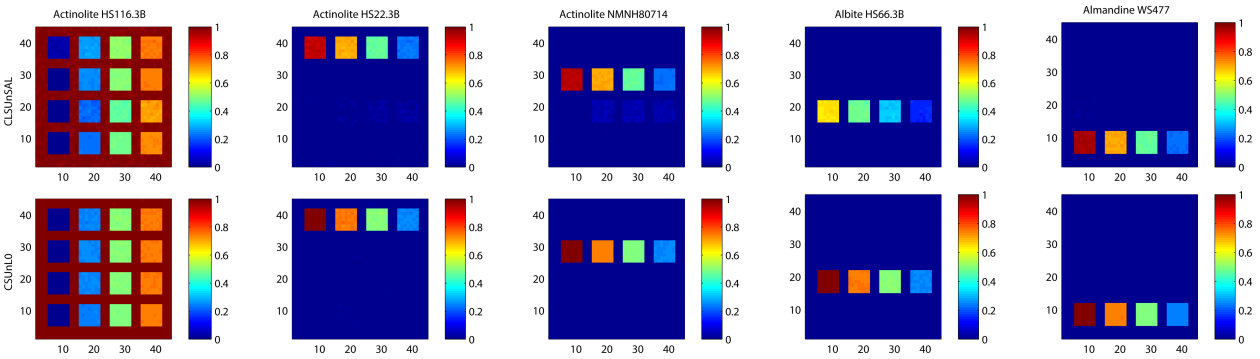

Fig. 7. Results of experiment 2: the SNR of white noise is 50 dB.

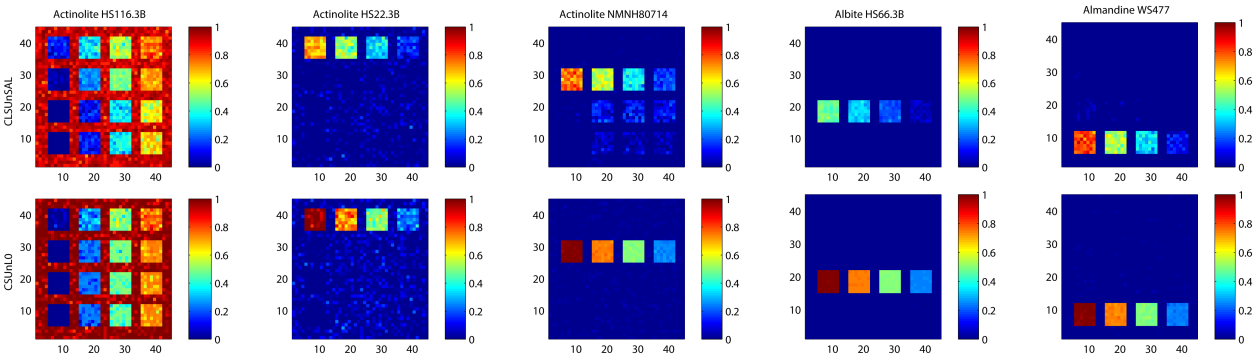

Fig. 8. Results of experiment 2: the SNR of white noise is 30 dB.

TABLE IV
RMSEs OBTAINED BY CLSUNSAL AND CSUNL0

RMSE(dB)	CLSUnSAL	CSUnL0
50	0.0285	0.0024
30	0.0851	0.0248

* $a_0 = 0.02$

Fig.7 and Fig.8 show that the results obtained by CLSUnSAL are smaller than CSUnL0's, which means the result of CLSUnSAL exists a lot of inactive endmembers to make abundances dispersive. On the contrary, CSUnL0 has no such problem, the row-hard-threshold function guarantees that only active endmembers are used to unmix the hyperspectal data in the iteration. Moreover, to compare between Fig.7 and Fig.8, CSUnL0's results are stable when the noise intensity increases. Thus, it can be believed that CSUnL0 has pruning dictionary ability and stronger anti-noise ability.

Now we discuss the parameter a in CSUnL0 which directly controls the result's sparsity. We still use the data set in this experiment with 30 dB noise, initialize $a_0 = 10^{-5}, 10^{-4.5}, ..., 10^{-1}$ respectively. Then we will compare CSUnL0's performance between these cases to probe the role of a, the results are shown as Table V and Fig.9.

Table V shows there is an interval for the parameter a to get the best result and achieve a suitable sparse degree, that is to say, an appropriate a exists and can make the algorithm find a good solution near the optimal point. From Table V

TABLE V
RMSEs AND SPARSITY WITH DIFFERENT PARAMETER VALUES

a_0	10^{-5}	$10^{-4.5}$	10^{-4}	$10^{-3.5}$	10^{-3}	$10^{-2.5}$	10^{-2}	$10^{-1.5}$	10^{-1}
RMSE	0.0960	0.0907	0.0759	0.0443	0.0246	0.0248	0.0249	0.0249	0.2195
$\|\|\mathbf{V}_2\|\|_{2,0}$	130	85	40	12	5	5	5	5	1

* $||\mathbf{V}_2||_{2,0}$ is the extractive endmembers' number, and the number of active endmembers is 5.

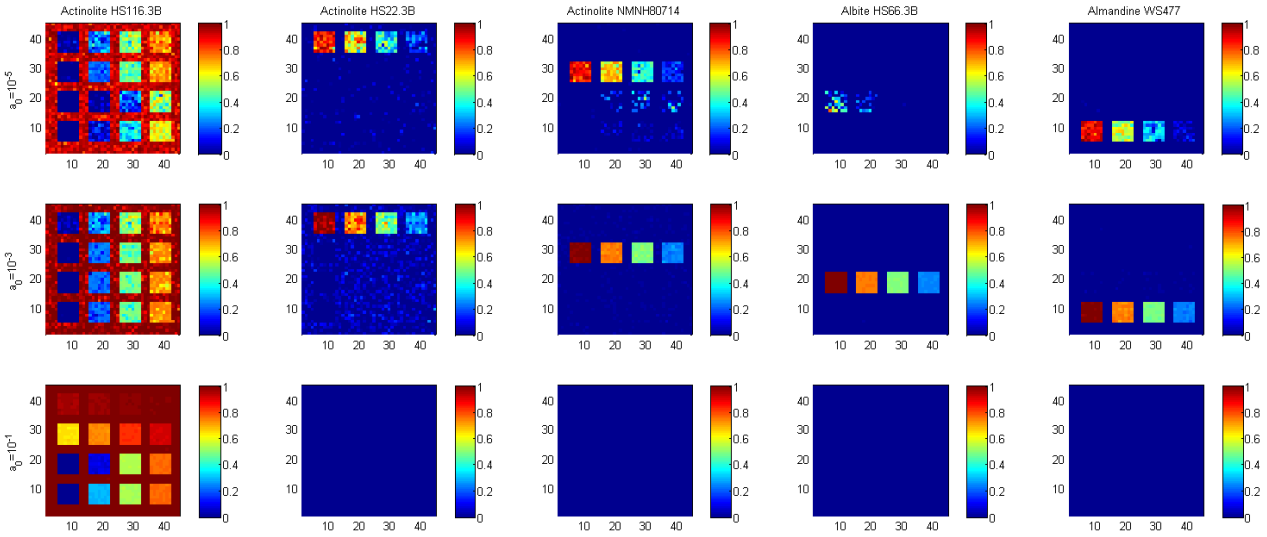

Fig. 9. Results of parameter experiment. From up to down, $a_0 = 10^{-5}, 10^{-3}, 10^{-1}$ respectively.

and Fig.9, if a_0 is very small, CSUnL0 can't exclude all inactive endmembers, so the result isn't sparse and accurate. On the contrary, if a_0 is very large, CSUnL0 will exclude some active endmembers by mistake. In other words, CSUnL0 can be tuned to the best performance, and the parameter a can control the result's sparsity.

D. Real hyperspectral data experiment

In some cases, simulation experiments cannot represent the real situation. So we use a 204×151 pixel subset of the well-known AVIRIS Cuprite data set to verify the effectiveness of CSUnL0. In this experiment, some wavebands of the low signal-to-noise ratio are removed, which means the spectral library $\mathbf{A}$ and the data set $\mathbf{Y}$ have only 188 wavebands. The results are shown in Fig.11. The real abundance distribution is the first column of Fig.11 and those pictures come from the software Tricorder 3.3[2] in 1995 which is shown in Fig.10, but the AVIRIS Cuprite data set was collected in 1997. Therefore, we can only regard this mineral classification figure as a reference to quantitatively evaluate different algorithms' performance.

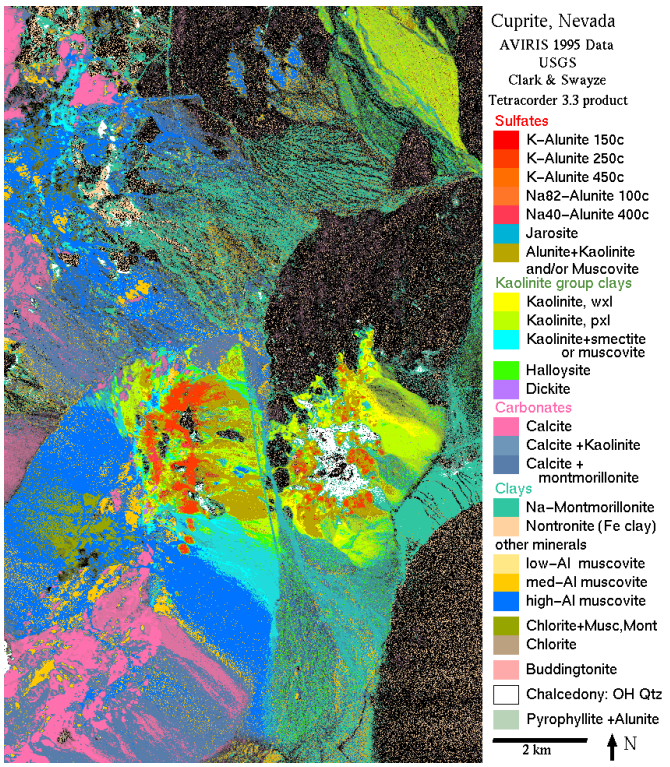

Fig. 10. USGS map showing the distribution of different minerals in the Cuprite mining district in Nevada.

[2]http://speclab.cr.usgs.gov/PAPERS/tetracorder/

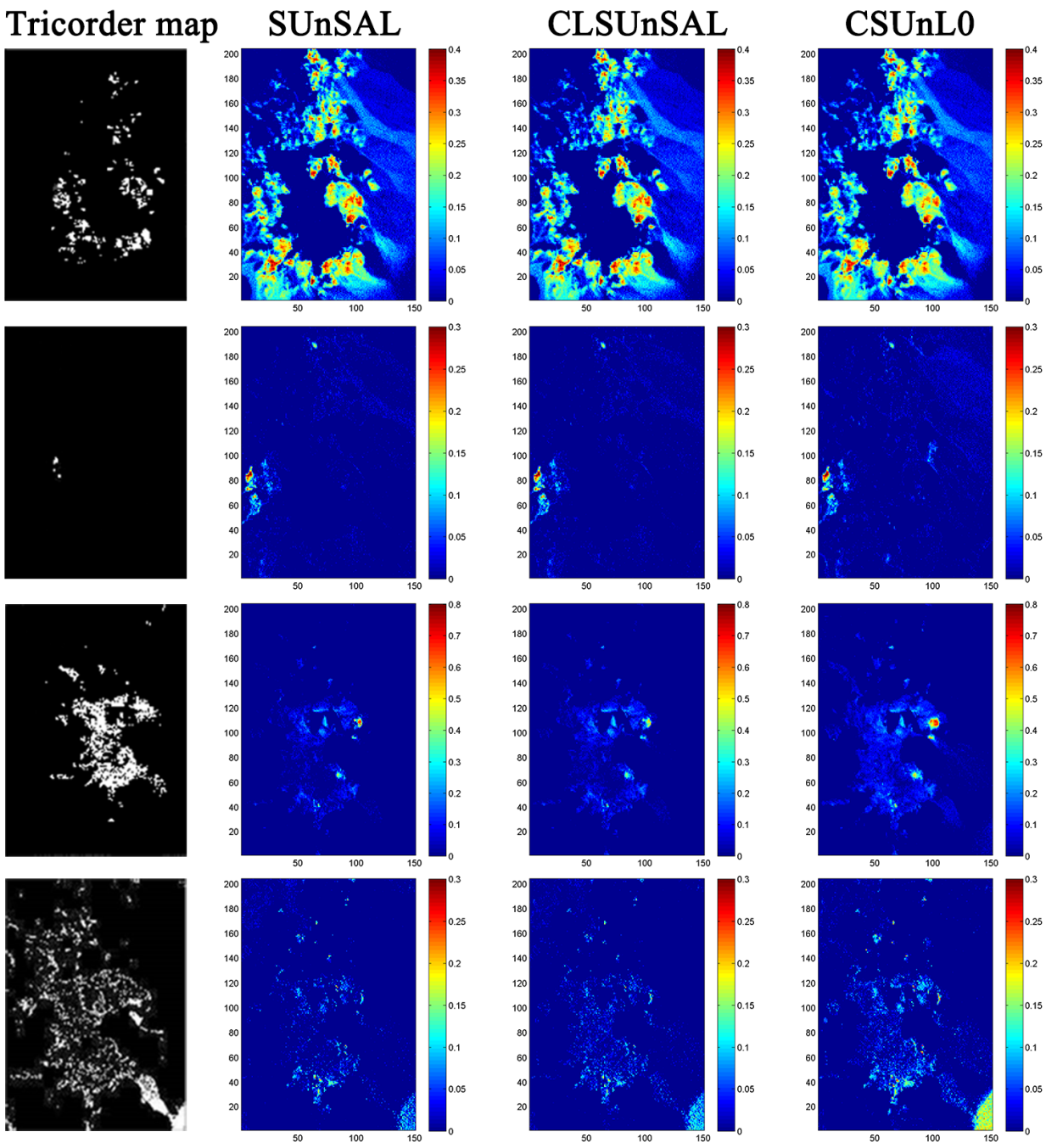

Fig. 11. The result of real hyperspectral data experiment. From up to down, maps correspond to Alunite, Buddingtonite, Chalcedony, and Montmorillonite respectively. From left to right, there are Tricorder map and abundances map estimated by SUnSAL, CLSUnSAL and CSUnL0. CSUnL0 parameter setting: $a_0 = 6 \times 10^{-5}$.

The Fig.11 shows that the results of algorithms are not exactly the same as Tricorder map. There are two reasons accounting for this phenomenon. On the one hand, when making Tricorder map, each pixel is assumed to consist of the only one endmember, which should certainly correspond to the endmember with the highest abundance in this pixel. However, when applying algorithms, we regard each pixel as a mixture of all spectra from different materials and assign abundances to these endmembers. On the other hand, in the spectral library, there are some materials having similar spectra. In Tricorder map, these materials with similar spectra are considered as the same, but they remain separated when using unmixing algorithms. Meanwhile, abundances of CSUnL0's result are always higher than the other two algorithms by excluding inactive endmembers. This will reduce the interference of noise. Thus, by solving l_0 norm problem, CSUnL0 can obtain more accurate results.

From all above, it can be claimed that CSUnL0 is a valid algorithm for the real data set.

V. Conclusion

In this paper, we propose a novel algorithm termed as CSUnL0 for the sparse hyperspectral unmixing problem. This algorithm is an improvement of CLSUnSAL which is an effective convex relaxation algorithm. According to the characteristics of materials distribution, CSUnL0 uses the $l_{2,0}$ norm constraint to replace the $l_{2,1}$ norm constraint in CLSUnSAL. In order to solve the NP-hard problem caused by l_0 norm constraint, this algorithm uses the variable splitting method and introduces a row-hard-threshold function to apply ADMM into the situation where l_0 norm exists. Especially, we present related convergence analysis for our algorithm. In this way, this NP-hard problem could be transformed into a new problem with polynomial complexity to find a good sub-optimal solution. Furthermore, the results of experiments with real and synthetic data sets indicate that the $l_{2,0}$ norm constraint contributes to getting a more accurate solution by excluding inactive endmembers. Meanwhile, the results also show that our proposed algorithm has a better anti-noise ability and robustness.

References

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[2] P. Shippert, "Why use hyperspectral imagery?" *Photogramm. Eng. Remote Sens.*, vol. 70, no. 4, pp. 377 – 380, Apr. 2004.

[3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17 – 28, Jan. 2002.

[4] M. Petrou and P. G. Foschi, "Confidence in linear spectral unmixing of single pixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1 pt 2, pp. 624 – 626, Jan. 1999.

[5] Y. Hu, H. Lee, and F. Scarpace, "Optimal linear spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1 pt 2, pp. 639 – 644, Jan. 1999.

[6] J. Boardman, "Automated spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1, 1993, pp. 11 – 14.

[7] M. E. Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 3753, Denver, CO, USA, Jul. 1999, pp. 266 – 275.

[8] J. M. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898 – 910, Apr. 2005.

[9] 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.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788 – 791, Oct. 1999.

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

[12] W. Tang, Z. Shi, and Z. An, "Nonnegative matrix factorization for hyperspectral unmixing using prior knowledge of spectral signatures," *Opt. Eng.*, vol. 51, no. 8, Aug. 2012.

[13] X. Chen, J. Chen, X. Jia, B. Somers, J. Wu, and P. Coppin, "A quantitative analysis of virtual endmembers' increased impact on the collinearity effect in spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2945 – 2956, Aug. 2011.

[14] M. Iordache, "A sparse regression approach to hyperspectral unmixing," Ph.D. dissertation, Univ. Tcnica Lisboa, Lisbon, Portugal, 2011.

[15] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6 PART 1, pp. 2014 – 2039, Jun. 2011.

[16] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 219 – 228, Jan. 2012.

[17] M. Elad, *Sparse and Redundant Representations*. Springer New York, 2010.

[18] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948 – 958, Jun. 2010.

[19] A. J. Miller, *Subset Selection in Regression*, 2nd ed. London: Chapman and Hall, 2002.

[20] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585 – 599, Feb. 2012.

[21] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. Cambridge, MA, USA: MIT Press, 2009.

[22] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397 – 3415, Dec. 1993.

[23] Z. Shi, W. Tang, Z. Duren, and Z. Jiang, "Subspace matching pursuit for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3256 – 3274, Jun. 2014.

[24] W. Tang, Z. Shi, and Y. Wu, "Regularized simultaneous forward-backward greedy algorithm for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5271 – 5288, Sep. 2014.

[25] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained $l_p - l_2$ optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142 – 1146, Sep. 2013.

[26] W. Tang, Z. Shi, and Z. Duren, "Sparse hyperspectral unmixing using an approximate l0 norm," *Optik*, vol. 125, no. 1, pp. 31 – 38, Jan. 2014.

[27] J. M. Bioucas-Dias and M. A. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," *Proc. IEEE WHISPERS*, Jun. 2010.

[28] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341 – 354, Jan. 2014.

[29] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4775 – 4789, Aug. 2016.

[30] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral a priori information," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 770 – 783, Feb. 2015.

[31] D. Wang, Z. Shi, and X. Cui, "Robust sparse unmixing for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1348 – 1359, 2018.

[32] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering Sparse Signals With a Certain Family of Nonconvex Penalties and DC Programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.

[33] A. J. Seneviratne, *l0 signal processing and model selection with applications*. New South Wales: University of New South Wales, 2013.

[34] X. Xu and Z. Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 54–69, Feb. 2017.

[35] E. Esser, "Applications of lagrangian-based alternating direction methods and connections to split bregman," *CAM report*, vol. 9, p. 31, 2009.

[36] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681 – 695, Mar. 2011.

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

[38] C. Ann Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2 II, pp. 1083 – 1094, Mar. 2000.

[39] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[40] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572 – 588, Mar. 2006.

[41] M. Kowalski and B. Torrésani, "Sparsity and persistence: Mixed norms provide simple signal models with dependent coefficients," *Signal, Image Video Process.*, vol. 3, no. 3, pp. 251 – 264, Sep. 2009.

[42] T. Blumensath and M. E. Davies, "Iterative Thresholding for Sparse Approximations," *J. Fourier Anal. Applicat.*, vol. 14, no. 5, pp. 629–654, Dec. 2008.

[43] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer-Verlag, 2006.
[44] R. N. Clark, G. A. Swayze, R. Wise, E. Livo, T. Hoefen, R. Kokaly, and S. J. Sutley, "Usgs digital spectral library splib06a." US Geol. Survey Denver, Denver, CO, USA, 2007.

APPENDIX

A. Proof of Lemma 1

Proof. Assume that the vectors $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$ are the ith row of the matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. Therefore, we can write the Lagrange function of the optimization problem as follows:

$$\begin{aligned} L(\boldsymbol{\Phi}) &= ||\boldsymbol{\Psi}-\boldsymbol{\Phi}||_F^2+\lambda||\boldsymbol{\Phi}||_{r-0} \\ &= \sum_i \left[\lambda||\boldsymbol{\phi}_i||_{r-0}+\sum_j(\Phi_{i,j}^2-2\Psi_{i,j}\Phi_{i,j})\right]+||\boldsymbol{\Psi}||_F^2 \\ &= \sum_i l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)+||\boldsymbol{\Psi}||_F^2 \end{aligned}$$

Ignore the constant term $||\boldsymbol{\Psi}||_F^2$, we can minimize $l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)=\lambda+\sum_j(\Phi_{i,j}^2-2\Psi_{i,j}\Phi_{i,j})$ as quadratic form by $\boldsymbol{\phi}_i=\boldsymbol{\psi}_i$ when $\boldsymbol{\phi}_i$ is a non-zero vector. Considering the situation $\boldsymbol{\phi}_i$ is zero vector, $l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)$ has two cases to achieve minimum:

$$l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)=\begin{cases} 0 & \boldsymbol{\phi}_i=\mathbf{0} \\ \lambda-\sum_j \Psi_{i,j}^2 & \boldsymbol{\phi}_i=\boldsymbol{\psi}_i \end{cases}$$

For each $l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)$, the second case can be positive when $\sum_j \Psi_{i,j}^2<\lambda$, and then $\boldsymbol{\phi}_i$ should be set to zero for minimum, which is just the operation of row-hard-threshold function. That is to say, the solution of $\min_{\boldsymbol{\phi}_i} l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)$ is $\boldsymbol{\phi}_i=\mathcal{RH}_\lambda(\boldsymbol{\psi}_i)$. Besides, $\sum_j \Psi_{i,j}^2=||\boldsymbol{\psi}_i||_2^2$ indicates the metric for a row in $\boldsymbol{\Psi}$ should be its l_2 norm.

Therefore, when each optimization subproblem $\min_{\boldsymbol{\phi}_i} l(\boldsymbol{\phi}_i,\boldsymbol{\psi}_i)$ is satisfied, the original optimization problem $\min_{\boldsymbol{\Phi}} L(\boldsymbol{\Phi})$ can be satisfied and the solution is:

$$\boldsymbol{\Phi}=\mathcal{RH}_\lambda(\boldsymbol{\Psi})$$

□

B. Proof of Theorem 1

Proof. As a whole, we focus on the monotonicity of Eq.(26) between k and $k+1$.

Firstly, the decline when updating $\mathbf{X}$ is:

$$\begin{aligned} &L(\mathbf{X}^k,\mathbf{V}_1^k,\mathbf{V}_2^k,\mathbf{D}_1^k,\mathbf{D}_2^k)-L(\mathbf{X}^{k+1},\mathbf{V}_1^k,\mathbf{V}_2^k,\mathbf{D}_1^k,\mathbf{D}_2^k) \\ &=\frac{\mu}{2}<\mathbf{X}^k-\mathbf{X}^{k+1},\mathbf{A}^T(\mathbf{A}\mathbf{X}^k+\mathbf{A}\mathbf{X}^{k+1}-2\mathbf{V}_1^k-2\mathbf{D}_1^k) \\ &\quad+\mathbf{X}^k+\mathbf{X}^{k+1}-2\mathbf{V}_2^k-2\mathbf{D}_2^k> \\ &=\frac{\mu}{2}<\mathbf{X}^k-\mathbf{X}^{k+1},(\mathbf{A}^T\mathbf{A}+\mathbf{I})(\mathbf{X}^k-\mathbf{X}^{k+1})> \\ &\geq\frac{\mu}{2}||\mathbf{X}^k-\mathbf{X}^{k+1}||_F^2 \end{aligned} \tag{31}$$

Secondly, define $g_1(\mathbf{V}_1)=\frac{1}{2}||\mathbf{Y}-\mathbf{V}_1||_F^2$. When updating $\mathbf{V}_1$, the dual feasible condition about $\mathbf{V}_1$ is true:

$$\begin{aligned} \mathbf{0}=\frac{\partial L}{\partial \mathbf{V}_1} &= \nabla g_1(\mathbf{V}_1^{k+1})+\mu(\mathbf{D}_1^k+\mathbf{V}_1^{k+1}-\mathbf{A}\mathbf{X}^{k+1}) \\ &= \nabla g_1(\mathbf{V}_1^{k+1})+\mu\mathbf{D}_1^{k+1} \\ &= \mathbf{V}_1^{k+1}-\mathbf{Y}+\mu\mathbf{D}_1^{k+1} \end{aligned} \tag{32}$$

Then after updating $\mathbf{V}_1$ and $\mathbf{D}_1$, the following inequality can be obtained by using the above dual feasible condition.

$$\begin{aligned} &L(\mathbf{X}^k,\mathbf{V}_1^k,\mathbf{V}_2^k,\mathbf{D}_1^k,\mathbf{D}_2^k)-L(\mathbf{X}^k,\mathbf{V}_1^{k+1},\mathbf{V}_2^k,\mathbf{D}_1^{k+1},\mathbf{D}_2^k) \\ &=g_1(\mathbf{V}_1^k)-g_1(\mathbf{V}_1^{k+1})+<\mu\mathbf{D}_1^{k+1},\mathbf{V}_1^k-\mathbf{V}_1^{k+1}>+ \\ &\quad\frac{\mu}{2}||\mathbf{V}_1^k-\mathbf{V}_1^{k+1}||_F^2-\frac{1}{\mu}||\mu\mathbf{D}_1^k-\mu\mathbf{D}_1^{k+1}||_F^2 \\ &=(\frac{1}{2}+\frac{\mu}{2}-\frac{1}{\mu})||\mathbf{V}_1^k-\mathbf{V}_1^{k+1}||_F^2 \\ &\geq\frac{\mu}{2}||\mathbf{V}_1^k-\mathbf{V}_1^{k+1}||_F^2 \\ &\text{subject to}:\mu\geq 2 \end{aligned} \tag{33}$$

Thirdly, assuming the row vectors $(\mathbf{v}_2)_i$, $(\mathbf{d}_2)_i$ and $(\mathbf{x})_i$ are the ith row in $\mathbf{V}_2$, $\mathbf{D}_2$ and $\mathbf{X}$, $a=\frac{2\lambda}{\mu}$, so the range of $||(\mathbf{v}_2)_i||_2^2$ calculated by row-hard-threshold function in Eq.(11) is:

$$||(\mathbf{v}_2)_i||_2^2\in\{0\}\cup(a,+\infty) \tag{34}$$

Then define $g_2(\mathbf{V}_2)=\lambda||\mathbf{V}_2||_{2,0}$ and the equivalent form $g_2((\mathbf{v}_2)_i)=\lambda\times 1(||(\mathbf{v}_2)_i||_2>0)\in\{0,\lambda\}$. When updating $\mathbf{V}_2$, the dual feasible condition about $\mathbf{V}_2$ only holds when $(\mathbf{v}_2)_i\neq\mathbf{0}$:

$$\begin{aligned} \mathbf{0}=\frac{\partial L}{\partial(\mathbf{v}_2)_i} &= \nabla g_2((\mathbf{v}_2^{k+1})_i)+\mu((\mathbf{d}_2^k)_i+(\mathbf{v}_2^{k+1})_i-(\mathbf{x}^{k+1})_i) \\ &= \nabla g_2((\mathbf{v}_2^{k+1})_i)+\mu(\mathbf{d}_2^{k+1})_i \\ &= \mu(\mathbf{d}_2^{k+1})_i \end{aligned} \tag{35}$$

Like the Eq.(33), the next equation can be gotten after updating $\mathbf{V}_2$ and $\mathbf{D}_2$.

$$\begin{aligned} &L(\mathbf{X}^k,\mathbf{V}_1^k,\mathbf{V}_2^k,\mathbf{D}_1^k,\mathbf{D}_2^k)-L(\mathbf{X}^k,\mathbf{V}_1^k,\mathbf{V}_2^{k+1},\mathbf{D}_1^k,\mathbf{D}_2^{k+1}) \\ &=g_2(\mathbf{V}_2^k)-g_2(\mathbf{V}_2^{k+1})+<\mu\mathbf{D}_2^{k+1},\mathbf{V}_2^k-\mathbf{V}_2^{k+1}>+ \\ &\quad\frac{\mu}{2}||\mathbf{V}_2^k-\mathbf{V}_2^{k+1}||_F^2-\mu||\mathbf{D}_2^k-\mathbf{D}_2^{k+1}||_F^2 \\ &=\sum_i\delta_i \end{aligned} \tag{36}$$

where

$$\begin{aligned} \delta_i=&g_2((\mathbf{v}_2^k)_i)-g_2((\mathbf{v}_2^{k+1})_i)+ \\ &<\mu(\mathbf{d}_2^{k+1})_i,(\mathbf{v}_2^k)_i-(\mathbf{v}_2^{k+1})_i>+ \\ &\frac{\mu}{2}||(\mathbf{v}_2^k)_i-(\mathbf{v}_2^{k+1})_i||_2^2-\mu||(\mathbf{d}_2^k)_i-(\mathbf{d}_2^{k+1})_i||_2^2 \end{aligned} \tag{37}$$

Besides, based on the row-hard-threshold function and dual variable updating formula in Note 1, we can elicit the following equations:

$$\begin{aligned}(\mathbf{v}_2^{k+1})_i &= \mathcal{RH}_a((\mathbf{x}^{k+1})_i - (\mathbf{d}_2^k)_i)\\ (\mathbf{d}_2^{k+1})_i &= (\mathbf{d}_2^k)_i - (\mathbf{x}^{k+1})_i + (\mathbf{v}_2^{k+1})_i\\ &= \begin{cases}\mathbf{0}, & ||(\mathbf{x}^{k+1})_i - (\mathbf{d}_2^k)_i||_2^2 > a\\ (\mathbf{d}_2^k)_i - (\mathbf{x}^{k+1})_i, & ||(\mathbf{x}^{k+1})_i - (\mathbf{d}_2^k)_i||_2^2 \le a\end{cases}\end{aligned} \tag{38}$$

That is to say $||(\mathbf{d}_2^{k+1})_i||_2^2 \le a$.

However, the dual feasible condition in Eq.(35) is not always true and $||(\mathbf{v}_2)_i||_2^2$ is in two ranges. So there are four cases which need be discussed when the states are k and $k+1$:

(1) If $(\mathbf{v}_2^k)_i = (\mathbf{v}_2^{k+1})_i = \mathbf{0}$, there is:

$$\begin{aligned}\delta_i &= -\mu||(\mathbf{d}_2^k)_i - (\mathbf{d}_2^{k+1})_i||_2^2\\ &\ge -\mu(||(\mathbf{d}_2^k)_i||_2 + ||(\mathbf{d}_2^{k+1})_i||_2)^2\\ &\ge -4\mu a = -8\lambda\end{aligned} \tag{39}$$

(2) If $||(\mathbf{v}_2^k)_i||_2^2 = 0, ||(\mathbf{v}_2^{k+1})_i||_2^2 > a$, then $||(\mathbf{d}_2^{k+1})_i||_2^2 = 0$, there is:

$$\delta_i = -\lambda + \frac{\mu}{2}||(\mathbf{v}_2^{k+1})_i||_2^2 - \mu||(\mathbf{d}_2^k)_i||_2^2 > -2\lambda \tag{40}$$

(3) If $||(\mathbf{v}_2^k)_i||_2^2 > a, ||(\mathbf{v}_2^{k+1})_i||_2^2 = 0$, then $||(\mathbf{d}_2^k)_i||_2^2 = 0$, there is:

$$\begin{aligned}\delta_i &= \lambda + < \mu(\mathbf{d}_2^{k+1})_i, (\mathbf{v}_2^k)_i > + \frac{\mu}{2}||(\mathbf{v}_2^k)_i||_2^2 - \mu||(\mathbf{d}_2^{k+1})_i||_2^2\\ &= \lambda + \frac{\mu}{2}||(\mathbf{d}_2^{k+1})_i + (\mathbf{v}_2^k)_i||_2^2 - \frac{3\mu}{2}||(\mathbf{d}_2^{k+1})_i||_2^2\\ &> -2\lambda\end{aligned} \tag{41}$$

(4) If $||(\mathbf{v}_2^k)_i||_2^2 > a, ||(\mathbf{v}_2^{k+1})_i||_2^2 > a$, then $||(\mathbf{d}_2^k)_i||_2^2 = ||(\mathbf{d}_2^{k+1})_i||_2^2 = 0$, there is:

$$\delta_i = \frac{\mu}{2}||(\mathbf{v}_2^k)_i - (\mathbf{v}_2^{k+1})_i||_2^2 \ge 0 \tag{42}$$

Therefore, based on above four cases, the following inequality is true:

$$\begin{aligned}&L(\mathbf{X}^k, \mathbf{V}_1^k, \mathbf{V}_2^k, \mathbf{D}_1^k, \mathbf{D}_2^k) - L(\mathbf{X}^k, \mathbf{V}_1^k, \mathbf{V}_2^{k+1}, \mathbf{D}_1^k, \mathbf{D}_2^{k+1})\\ &\qquad \ge -8m\lambda\end{aligned} \tag{43}$$

where m is the number of rows in $\mathbf{V}_2$.

As a conclusion, the change of the original Lagrange function during an iteration cycle can be expressed as follow:

$$\begin{aligned}&\Delta =\\ &L(\mathbf{X}^k, \mathbf{V}_1^k, \mathbf{V}_2^k, \mathbf{D}_1^k, \mathbf{D}_2^k) - L(\mathbf{X}^{k+1}, \mathbf{V}_1^{k+1}, \mathbf{V}_2^{k+1}, \mathbf{D}_1^{k+1}, \mathbf{D}_2^{k+1})\\ &\ge \frac{\mu}{2}||\mathbf{X}^k - \mathbf{X}^{k+1}||_F^2 + \frac{\mu}{2}||\mathbf{V}_1^k - \mathbf{V}_1^{k+1}||_F^2 - 8m\lambda\\ &\text{subject to}: \mu \ge 2\end{aligned} \tag{44}$$

That is to say, when $\lambda \le \frac{\mu}{16m}(||\mathbf{X}^k - \mathbf{X}^{k+1}||_F^2 + ||\mathbf{V}_1^k - \mathbf{V}_1^{k+1}||_F^2)$ and $\mu \ge 2$, the Lagrange function in Eq.(26) is decreasing.

Furthermore, the above formulas only give a dynamic range for λ, then we try to establish the connection between parameter λ and normalized error tolerance ϵ in Note 1 for getting a universal and fixed parameter. In other words, there is always a parameter λ to ensure the decreasing property when giving any of normalized error tolerance ϵ.

Firstly, we relax the following parameter and variable:

(1) Shrink the upper bound of λ.

Obviously, $\frac{\mu}{2}||\mathbf{X}^k - \mathbf{X}^{k+1}||_F^2 > 0$ is always true, so we ignore this item in Eq.(44) and $\lambda \le \frac{\mu}{16m}||\mathbf{V}_1^k - \mathbf{V}_1^{k+1}||_F^2$ can be obtained.

Then based on Eq.(32), there is $||\mathbf{V}_1^k - \mathbf{V}_1^{k+1}||_F^2 = \mu^2||\mathbf{D}_1^k - \mathbf{D}_1^{k+1}||_F^2$. So the following inequality can also promise that Eq.(44) is always non-negative:

$$\lambda \le \frac{\mu^3}{16m}||\mathbf{D}_1^k - \mathbf{D}_1^{k+1}||_F^2 \tag{45}$$

(2) Magnify the value of $||\mathbf{D}_2^k - \mathbf{D}_2^{k+1}||_F$.

Based on Eq.(38), there is:

$$||\mathbf{D}_2||_F = \sqrt{\sum_i^m ||(\mathbf{d}_2)_i||_2^2} \le \sqrt{ma} = \sqrt{\frac{2m\lambda}{\mu}} \tag{46}$$

Therefore, the following inequality can be obtained by using the triangle inequality:

$$||\mathbf{D}_2^k - \mathbf{D}_2^{k+1}||_F \le ||\mathbf{D}_2^k||_F + ||\mathbf{D}_2^{k+1}||_F \le 2\sqrt{\frac{2m\lambda}{\mu}} \tag{47}$$

Secondly, assume that the algorithm reaches the stopping criterion when $k = k_s > 0$. The following inequality can be derived based on Note 1:

$$\begin{aligned}&||\mathbf{V}_1^{k_s+1} - \mathbf{A}\mathbf{X}^{k_s+1}||_F + ||\mathbf{V}_2^{k_s+1} - \mathbf{X}^{k_s+1}||_F\\ &\quad = ||\mathbf{D}_1^{k_s+1} - \mathbf{D}_1^{k_s}||_F + ||\mathbf{D}_2^{k_s+1} - \mathbf{D}_2^{k_s}||_F\\ &\quad < \sqrt{(m+L)\times K}\epsilon\end{aligned} \tag{48}$$

Then based on Eq.(47) and Eq.(48), for all $k < k_s$, there is:

$$||\mathbf{D}_1^{k+1} - \mathbf{D}_1^k||_F \ge \sqrt{(m+L)\times K}\epsilon - 2\sqrt{\frac{2m\lambda}{\mu}} \tag{49}$$

Finally, when putting Eq.(49) into Eq.(45), we can obtain a universal range for λ to guarantee the effectiveness of our algorithm before reaching the stopping criterion. The range is as follow:

$$\lambda \le \frac{\mu^3}{16m}\left(\sqrt{(m+L)\times K}\epsilon - 2\sqrt{\frac{2m\lambda}{\mu}}\right)^2 \tag{50}$$

$$\Rightarrow \quad \lambda \le \frac{\mu(m+L)K\epsilon^2}{m(\frac{4}{\mu} + 2\sqrt{2})^2} \tag{51}$$

Besides, based on Eq.(32), Eq.(46) and $\mu > 2$, the lower bound of the augmented Lagrange function can be found as follow:

$$\begin{aligned}&L(\mathbf{X}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{D}_1, \mathbf{D}_2)\\ &\quad \ge \frac{1}{2}||\mathbf{Y} - \mathbf{V}_1||_F^2 - \frac{\mu}{2}||\mathbf{D}_1||_F^2 - \frac{\mu}{2}||\mathbf{D}_2||_F^2\\ &\quad \ge \frac{1}{2}(1 - \frac{1}{\mu})||\mathbf{Y} - \mathbf{V}_1||_F^2 - \sqrt{\frac{m\mu\lambda}{2}} \ge -\sqrt{\frac{m\mu\lambda}{2}}\end{aligned} \tag{52}$$

□