Subspace Matching Pursuit for Sparse Unmixing of Hyperspectral Data

Zhenwei Shi, *Member, IEEE,* Wei Tang, Zhana Duren, Zhiguo Jiang

Abstract

Sparse unmixing assumes that each mixed pixel in the hyperspectral image can be expressed as a linear combination of only a few spectra (endmembers) in a spectral library, known a priori. It then aims at estimating the fractional abundances of these endmembers in the scene. Unfortunately, because of the usually high correlation of the spectral library, sparse unmixing problem still remains a great challenge. Besides, most related work focuses on the l_1 convex relaxation methods and little attention has been paid to the use of simultaneous sparse representation via greedy algorithms (SGA) for sparse unmixing. SGA has advantages such as it can get an approximate solution for the l_0 problem directly without smoothing the penalty term in a low computational complexity as well as exploit the spatial information of the hyperspectral data. Thus, it is necessary to explore the potential of using such algorithms for sparse unmixing. Inspired by the existing SGA methods, this paper presents a novel greedy algorithm termed subspace matching pursuit (SMP) for sparse unmixing of hyperspectral data. SMP makes use of the low-degree mixed pixels in the hyperspectral image to iteratively find a subspace to reconstruct the hyperspectral data. It is proved that under certain conditions, SMP can recover the optimal endmembers from the spectral library. Moreover, SMP can serve as a dictionary pruning algorithm. Thus, it can boost other sparse unmixing algorithms, making them more accurate and time-efficient. Experimental results on both synthetic and real data demonstrate the efficacy of the proposed algorithm.

Index Terms

Hyperspectral unmixing, sparse unmixing, dictionary pruning, simultaneous sparse representation, multiple-measurement vector (MMV), greedy algorithm (GA), subspace matching pursuit (SMP).

Zhenwei Shi, Wei Tang and Zhiguo Jiang are with Image Processing Center, School of Astronautics, Beihang University, Beijing, P.R. China, 100191 (E-mail: shizhenwei@buaa.edu.cn (Zhenwei Shi); tangwei@sa.buaa.edu.cn (Wei Tang); jiangzg@buaa.edu.cn (Zhiguo Jiang)). Zhana Duren is with National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (E-mail: durenzn@amss.ac.cn).

The work was supported by the National Natural Science Foundation of China under the Grants 61273245, 61071137 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.

I. INTRODUCTION

Hyperspectral remote sensing measures radiance from earth's surface materials at hundreds of narrow and contiguous wavelength bands. During recent years, hyperspectral sensors have been evolved to collect spectra extending from visible region through the infrared band ($0.3\ \mu m - 2.5\ \mu m$). Hyperspectral remote sensing has been widely applied in the fields of target detection, material mapping, material identification and mapping details of surface properties [1, 2]. Frequently, however, due to the low spatial resolution of a sensor as well as the combination of distinct materials into a homogeneous mixture, each pixel in the hyperspectral image often contains more than one pure substance [3]. Unmixing, a great challenging task underlying many hyperspectral image applications, then aims at decomposing the measured spectrum of each mixed pixel into a collection of constituent spectra (endmembers) and a set of corresponding fractions (abundances).

Depending on the mixing scales at each pixel and on the geometry of the scene, the observed mixture is either linear or nonlinear. For the past few years, linear unmixing model has been widely used to solve the unmixing problem for its advantages such as ease of implementations and flexibility in different applications [4–6]. It considers each mixed pixel a linear combination of endmembers weighted by their corresponding abundance fractions. Under this model, a group of unmixing approaches based on geometry [7–14], statistics [15, 16] and nonnegative matrix factorization (NMF) [17–23] have been proposed. However, all these methods are unsupervised and could extract virtual endmembers [24] with no physical meaning. To overcome this difficulty, a semi-supervised unmixing approach has been proposed to model each mixed pixel in the hyperspectral image using only a few spectra in a spectral library, known a priori. Because the number of spectral signatures contained in the spectral library is much larger than the number of endmembers present in the hyperspectral image, this approach often leads to a sparse solution [25]. The newly formulated problem (called sparse unmixing) is combinatorial and needs efficient sparse regression techniques to solve it [6].

Several sparse regression techniques have been used for sparse unmixing in [25], including orthogonal matching pursuit (OMP), basis pursuit (BP), BP denoising and iterative spectral mixture analysis (ISMA). In [26], sparse unmixing is tackled from a Bayesian perspective. Recently, some sparse unmixing methods have been proposed that exploit the contextual information [28, 29] and subspace nature [30, 31] of the hyperspectral image to obtain a much better result. As this approach 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 identify such pure signatures.

To the best of our knowledge, most sparse unmixing methods are based on convex relaxation methods; little attention has been paid to the use of greedy algorithms for sparse unmixing. Though OMP, a typical greedy algorithm, is used to solve the unmixing problem in [25], the high correlation of the spectral library blocks it from getting the optimal solution. On the other hand, the advantages brought by greedy algorithms such as low computational complexity of solving the optimization problems containing non-smooth terms and that they can get an approximate solution for the l_0 problem directly without smoothing the penalty function are attractive. Therefore

it is necessary to explore the potential of greedy algorithms in sparse unmixing. Under this circumstance, we consider using the simultaneous sparse representation via greedy algorithms (SGA) for sparse unmixing. Some SGA methods have been used for hyperspectral detection [45] and hyperspectral classification [27]. SGA which assumes that the input data are constructed with the same set of signatures can exploit the spatial information of the hyperspectral data. This characteristic of SGA can alleviate the sparse regression limitations linked to the usually high correlation of the hyperspectral libraries which is the major challenging problem to be tackled for the sparse unmixing algorithms. However, one obstacle faced by such algorithms is that they tend to be trapped into local optimum and are likely to miss some of the actual endmembers when the number of endmembers present in the hyperspectral scene is large. Thus, we develop a block-processing strategy to sidestep this obstacle.

Inspired by the existing SGA methods, we propose a novel sparse unmixing algorithm termed subspace matching pursuit (SMP) in this paper. SMP exploits the fact that the spectral vectors in the hyperspectral image are highly correlated and some endmembers present in one mixed pixel are also probably present in the others. It iteratively selects members from the spectral library to reconstruct the whole hyperspectral data. As an SGA method, SMP can make use of the spatial information within the hyperspectral image. Besides, since SMP utilizes the low-degree mixed pixels in the hyperspectral data, it can find the actual endmembers more accurately than the other considered SGA methods. It also has the advantage of low computational complexity and implementing very fast. Furthermore, it can serve as a dictionary pruning algorithm and boost other sparse unmixing algorithms, making them more time-efficient and accurate. Theorems are provided to analyse the conditions under which SMP will recover the optimal endmembers present in the hyperspectral image from the spectral library. To evaluate the performance, the method is tested on synthetic and real hyperspectral data. The results of those experiments demonstrate its effectiveness.

The rest of the paper is structured as follows. Section II introduces the sparse unmixing problem and briefly reviews some algorithms for sparse unmixing. In Section III, we introduce the simultaneous sparse representation via greedy algorithms for sparse unmixing. In Section IV we present the proposed subspace matching pursuit algorithm and give out some theoretical analyses for the algorithm. Experimental results are shown in Section V. Finally, we conclude in Section VI.

II. Sparse Unmixing of Hyperspectral Data

In this section, we first review the widely used linear mixture model. Then, the sparse unmixing problem is presented and some algorithms for sparse unmixing are introduced at the same time.

The linear sparse unmixing model assumes that the observed spectrum of a mixed pixel can be expressed as a linear combination of only a few spectral signatures selected from a library known in advance. Suppose $\mathbf{A} \in R^{L\times m}$ is the spectral library, where L is the number of bands and m is the number of spectral signatures in the library. Then, the sparse unmixing model can be written as follows:

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

where $\mathbf{y} \in R^L$ is the spectrum vector of a mixed pixel with L bands, $\mathbf{x} \in R^m$ is the abundance vector with regard to the library $\mathbf{A}$, $\mathbf{n} \in R^L$ is the vector of error term. Owing to physical constraints, the fractional abundances of

all the endmembers for a mixed pixel can not be negative and their sum should be one. Then the model has the following two constraints (x_i is the ith element of $\mathbf{x}$):

$$x_i \geq 0, \forall i \tag{2}$$

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

which are called nonnegativity and sum-to-one, respectively. As the number of endmembers involved in a mixed pixel is usually very small compared with the dimensionality of the spectral library, the vector of fractional abundances $\mathbf{x}$ is sparse.

Without considering the two constraints in Eq. (2) and Eq. (3), the optimization problem of sparse unmixing is then

$$(P_0): \quad \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \delta \tag{4}$$

where $\|\mathbf{x}\|_0$ (called l_0 norm) denotes the number of nonzero components in $\mathbf{x}$, $\delta \geq 0$ is the error tolerance due to the noise and modeling errors. However, problem (P_0) is an NP-hard problem which means it is very combinatorial and complex to solve. Three kinds of methods can tackle this problem, namely greedy algorithms, convex relaxation methods and sparse Bayesian methods.

For sparse unmixing, problem (P_0) can be divided into two questions: i) which spectral signatures should be selected from the spectral library $\mathbf{A}$ to model the spectral vector of the pixel $\mathbf{y}$; ii) what are the abundances for these selected signatures. The second question can be easily solved by the least squares techniques once the first question has been solved. As the support of the solution $\mathbf{x}$ is discrete in nature, algorithms that seek it are discrete as well [35]. This line of reasoning makes us turn to the family of greedy algorithms for the solving of this problem.

Greedy algorithms follow the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum [36]. Thus, a greedy strategy could not produce an optimal solution for some problems, nonetheless it is able to yield locally optimal solutions that approximate a global optimal solution in reasonable time. Orthogonal matching pursuit (OMP) [37], a typical greedy algorithm, has been utilized to solve (P_0) in [25]. OMP iteratively selects at each step the member in the spectral library best correlated with the residual part of the mixed pixel $\mathbf{y} - \mathbf{A}\mathbf{x}$. Then it produces a new approximation by projecting the spectral vector of the mixed pixel onto the potential endmembers that have already been selected. The algorithm stops when the condition $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \delta$ is satisfied. A member from the spectral library can not be selected more than once as the residual is orthogonalized with respect to the members already selected.

When taking the constraint of nonnegativity in Eq. (2) into account, problem (P_0) transforms into

$$(P_0^+): \quad \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \delta, \mathbf{x} \geq 0. \tag{5}$$

A variant of OMP has been proposed in [38] to solve problem (P_0^+). It is worth mentioning that we do not explicitly add the sum-to-one constraint in Eq. (3) 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 [25][41].

Another approach to get the sparse solution is to relax the l_0 norm to any continuous function $\mathbf{J}(\mathbf{x})$ that favours sparsity

$$(P_J^+): \quad \min_{\mathbf{x}} \mathbf{J}(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \leq \delta, \mathbf{x} \geq 0. \tag{6}$$

Common choices of $\mathbf{J}(\mathbf{x})$ can be i) the l_p "norms"[1] with $p < 1$: $\|\mathbf{x}\|_p^p = \sum_i |x_i|^p$, ii) the l_1 norm, iii) other functions $\mathbf{J}(\mathbf{x}) = \sum_i \rho(x_i)$ that promotes sparsity, such as $\rho(x_i) = 1 - \exp(|x_i|)$, $\rho(x_i) = \log(1 + |x_i|)$, and $\rho(x_i) = |x_i|/(1 + |x_i|)$ [35]. Among them the l_1 convex relaxation methods (i.e. $\mathbf{J}(\mathbf{x}) = \|\mathbf{x}\|_1$) have been widely researched. Then problem (P_0^+) turns into the following optimization problem, known in the literature as constrained basis pursuit denoising (CBPDN) [39]:

$$(P_1^+): \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \leq \delta, \mathbf{x} \geq 0. \tag{7}$$

As the l_1 norm is continuous and convex, problem (P_1^+) is more tractable than problem (P_0^+). Recently, the constrained spectral unmixing by variable splitting and augmented Lagrangian (CSUnSAL) [42] has been proposed to solve (P_1^+). CSUnSAL exploits the alternating direction method of multipliers (ADMM), which decomposes a difficult problem into a sequence of simpler ones.

For an appropriate Lagrange multiplier λ, the solution to problem (P_1^+) is precisely the solution to the problem of the constrained Lasso [40]:

$$(P_1^{\lambda+}): \quad \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{x} \geq 0 \tag{8}$$

where $\lambda > 0$ is the Lagrange multiplier and larger λ leads to sparser solution. Similar as CSUnSAL, spectral unmixing by variable splitting and augmented Lagrangian (SUnSAL) [42], which is also based on ADMM, can solve problem $(P_1^{\lambda+})$.

Recently, efforts have been paid to exploit the spatial-contextual information present in the hyperspectral images. Specifically, a total variation regularizer was added into problem (P_1^+). This idea leads to the sparse unmixing via variable splitting augmented Lagrangian and total variation (SUnSAL-TV) [29] and alternating direction method for deblurring and unmixing of hyperspectral images [28].

Convex relaxation methods are far more sophisticated than the greedy algorithms as they obtain the global solution of a well-defined optimization problem [35]. However, this characteristic of convex relaxation methods also means that they are far more complicated than the greedy algorithms.

The sparse Bayesian methods [34] have been applied to the sparse unmixing problem recently. In [26], sparse unmixing is formulated as a hierarchical Bayesian inference problem; priors for the model parameters are selected to ensure the nonnegativity and sparsity of the abundances' vector. Then an iterative scheme termed Bayesian inference iterative conditional expectations (BI-ICE) is proposed to estimate the model parameters. The BI-ICE algorithm can provide the sparse solution without necessarily tuning any parameters. However, it is much more complex than the convex relaxation methods and greedy algorithms.

[1]The l_p "norms" ($p < 1$) are no longer formal norms, as the triangle inequality is no longer satisfied.

III. Simultaneous Sparse Representation Via Greedy Algorithms for Sparse Unmixing

In this section, we first describe the sparse unmixing model that the simultaneous sparse representation via greedy algorithms (SGA) tackles. Then, two representative SGA methods, simultaneous orthogonal matching pursuit (SOMP) and simultaneous subspace pursuit (SSP), are extended to solve the unmixing problem.

SGA can obtain an approximation of several input signals using different linear combinations of the same elementary signals. As the sparse unmixing problem can be considered as using different linear combinations of the potential endmembers selected from the spectral library to reconstruct the input hyperspectral signatures, we can use SGA methods for sparse unmixing. Then the sparse unmixing model in Eq. (1) becomes

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{9}$$

where $\mathbf{Y} \in R^{L\times K}$ is the hyperspectral data matrix with L bands and K mixed pixels, $\mathbf{A} \in R^{L\times m}$ is the spectral library, $\mathbf{X} \in R^{m\times K}$ is the abundance matrix each column of which corresponds with the abundance fractions of a mixed pixel and $\mathbf{N} \in R^{L\times K}$ is the matrix of error term.

Under this model, the unmixing problem is as follows:

$$(P_{SGA}): \quad \min_{\mathbf{X}} \|\mathbf{X}\|_{\text{row-0}} \quad \text{subject to} \quad \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F \leq \delta \tag{10}$$

where $\|\mathbf{A}\|_F$ means the Frobenius norm[2] of matrix $\mathbf{A}$, $\|\mathbf{X}\|_{\text{row-0}}$ (called row-l_0 quasi-norm [43]) is the number of nonzero rows in matrix $\mathbf{X}$, δ is the error tolerance due to the noise and modeling errors. In [30], the authors relax problem (P_{SGA}) to a convex problem and develop a collaborative sparse unmixing method to solve it.

We note that the model in Eq. (10) is reasonable because a hyperspectral image always contains a small number of endmembers (especially compared with the very large spectral library) which means that there should be only a few rows with nonzero entries in the abundance matrix $\mathbf{X}$ [30]. Furthermore, this model has several advantages over the model in Eq. (4): i) as researched in the literature of multiple-measurement vectors (MMV) [44], the model in Eq. (10) is more likely to have a unique solution compared with the model in Eq. (4), which can also be seen in the upcoming theorem (i.e. Theorem 3.1); ii) the constraint that all the pixels in a hyperspectral image share the same active set of endmembers can alleviate the sparse regression limitations caused by the usually high correlation of the spectral library [30]; iii) most of the time, sparse regression via simultaneous sparse representation is superior to the non-simultaneous one [32]. Nevertheless, there are also some cases in which the abundance matrix $\mathbf{X}$ could be constructed such that it has no zero row but each pixel adopts a sparse representation in $\mathbf{X}$, especially when the hyperspectral image is very complex. In this paper, we only focus on the situations in which the number of endmembers present in the hyperspectral image is much smaller than the number of spectral signatures contained in the spectral library.

Here, we review a theorem that guarantees the uniqueness of the sparse representation of problem (P_{SGA}) in the noiseless case. This theorem will help to analyze the problem in the ideal situation. The following theorem was first introduced and proved in [44]:

[2] $\|\mathbf{A}\|_F = (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}}$, a_{ij} is the element of $\mathbf{A}$ that has coordinate (i, j).

Theorem 3.1: Matrix $\mathbf{X}$ will be the unique solution to the problem

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{\text{row-0}} \quad \text{subject to} \quad \mathbf{Y} = \mathbf{A}\mathbf{X} \tag{11}$$

if the following inequality strictly holds

$$\|\mathbf{X}\|_{\text{row-0}} < \frac{[Spark(\mathbf{A}) - 1 + Rank(\mathbf{Y})]}{2} \tag{12}$$

where $Spark(\mathbf{A})$ (or σ) is the smallest possible integer such that there exist σ columns of matrix $\mathbf{A}$ that are linearly dependent.

Even though $Spark(\mathbf{A})$ is very difficult to obtain (as it calls for a combinatorial search over all possible subsets of columns from $\mathbf{A}$), the upper bound in Theorem 3.1 is optimistic. As the number of endmembers is usually much smaller than the number of bands and number of pixels in a hyperspectral image, we can expect that the sparse representation of problem (P_{SGA}) in the noiseless case is unique.

Now, we introduce the SGA methods into sparse unmixing of hyperspectral data. Two algorithms are considered in this paper, namely simultaneous orthogonal matching pursuit (SOMP) and simultaneous subspace pursuit (SSP). These two algorithms have been used for hyperspectral target detection [45] and hyperspectral classification [27].

SOMP [43] is an extension of OMP. In the greedy selection step, instead of picking the member most correlated with the current residual of one input signal, SOMP chooses the member that maximizes the sum of absolute correlation with the residuals of all the input signals. In this way, SOMP can find a member from the library that has a strong overall contribution to the image. When applied to the sparse unmixing, SOMP can exploit the spatial information of the hyperspectral data and thus alleviate the sparse regression limitations caused by the usually high correlation of the spectral library. However, one great challenge faced by SOMP is that when the number of endmembers present in the scene is large, it tends to be trapped into the local optimum and miss some of the actual endmembers. Thus, we develop a block-processing strategy to sidestep this obstacle. Specifically, we divide the whole hyperspectral image into several blocks. Then in each block SOMP will pick several potential endmembers from the spectral library. Finally, the abundances are computed using the whole hyperspectral data and potential endmembers picked associated with all the blocks. Note this strategy is different with that in [45]. As the abundances are estimated using potential endmembers selected from all the blocks, our strategy can use the spatial information further.

The intuition behind this strategy is that i) the number of endmembers in one block is likely to be less than that in the whole scene and SOMP tends to be more effective when the number of members that construct the input signals is not large [43]; ii) by processing block by block, SOMP can utilize the contextual spatial information within the block; iii) one endmember present in one mixed pixel is probably also present in the others (either neighborly or non-neighborly) of the whole image and we can use the potential endmembers picked from all the blocks to reconstruct the hyperspectral image.

Besides, as a preprocessing step, we successively apply the zero-mean normalization and l_2 unit-length normalization to the columns of both the hyperspectral data matrix and spectral library matrix. To get a more effective

Algorithm 1: Simultaneous orthogonal matching pursuit for hyperspectral sparse unmixing

1: Apply zero-mean normalization and l_2 unit-length normalization successively to the columns of hyperspectral data matrix $\mathbf{Y}$ and spectral library matrix $\mathbf{A}$
2: Divide the hyperspectral image into several blocks, initialize the index set $\mathbf{S} = \emptyset$
3: **For** each block **do**
4: Set index set $\mathbf{S}_b = \emptyset$, iteration counter $k = 1$. Initialize the residual matrix $\mathbf{R}^0 = \mathbf{Y}_b$; $\{\mathbf{Y}_b$ is the submatrix of $\mathbf{Y}$ corresponding to the block$\}$
5: **While** stopping criterion has not been met **do**
6: $j = \arg\max_i \|(\mathbf{R}^{k-1})^T \mathbf{A}_i\|_2$; $\{\mathbf{A}_i$ is the ith column of $\mathbf{A}\}$
7: Set $\mathbf{S}_b = \mathbf{S}_b \bigcup j$
8: Compute $\mathbf{P} = (\mathbf{A}_{\mathbf{S}_b}^T \mathbf{A}_{\mathbf{S}_b})^{-1} \mathbf{A}_{\mathbf{S}_b}^T \mathbf{Y}_b$; $\{\mathbf{A}_{\mathbf{S}_b}$ is the matrix containing the columns of $\mathbf{A}$ having the indices from $\mathbf{S}_b\}$
9: Calculate the residual: $\mathbf{R}^k = \mathbf{Y}_b - \mathbf{A}_{\mathbf{S}_b} \mathbf{P}$
10: $k \leftarrow k + 1$
11: **End while**
12: Set $\mathbf{S} = \mathbf{S} \bigcup \mathbf{S}_b$
13: **End for**
14: Estimate abundances using the original hyperspectral data matrix and spectral library matrix:
$\mathbf{X} \leftarrow \arg\min_{\mathbf{X}} \|\mathbf{A}_{\mathbf{S}}\mathbf{X} - \mathbf{Y}\|$, subject to $\mathbf{X} \geq 0$

estimation of the abundances, the abundance nonnegativity constraint is added in. The whole process of using SOMP for sparse unmixing of hyperspectral data is summarized in Algorithm 1.

Simultaneous subspace pursuit (SSP) is an extension of subspace pursuit (SP) [46]. SSP is first proposed for hyperspectral target detection [45] and hyperspectral classification [27]. Similar to SP, SSP also maintains a list of c potential members. In the greedy selection step, the c members that get the best simultaneous representation for the residuals of all the input signals are selected as the new candidates. We also modify the algorithm in the same way as SOMP to improve its performance in sparse unmixing. The SSP algorithm for sparse unmixing is shown in Algorithm 2.

IV. Subspace Matching Pursuit (SMP)

In this section, we first present our new algorithm, subspace matching pursuit (SMP), for sparse unmixing of hyperspectral data. Then we analyze that SMP can also serve as a dictionary pruning algorithm to boost other sparse unmixing algorithms. Finally, some theoretical analyses of the algorithm will be given.

A. Statement of Algorithm

SMP belongs to the class of SGA method and it is an iterative technique. It is proposed to get an approximate solution for the problem (P_{SGA}) with the abundance nonnegativity constraint. The whole process of subspace

Algorithm 2 Simultaneous subspace pursuit for hyperspectral sparse unmixing

1: Apply zero-mean normalization and l_2 unit-length normalization successively to the columns of hyperspectral data matrix $\mathbf{Y}$ and spectral library matrix $\mathbf{A}$

2: Divide the hyperspectral image into several blocks, initialize the index set $\mathbf{S} = \emptyset$

3: **For** each block **do**

4: Set index set $\mathbf{S}_b = \{c$ indices corresponding to the c largest numbers in $\|\mathbf{Y}_b^T \mathbf{A}_i\|_2, i = 1, \cdots, m\}$, iteration counter $k = 1$. Initialize the residual matrix $\mathbf{R}^0 = \mathbf{Y}_b - \mathbf{A}_{\mathbf{S}_b}(\mathbf{A}_{\mathbf{S}_b}^T \mathbf{A}_{\mathbf{S}_b})^{-1} \mathbf{A}_{\mathbf{S}_b}^T \mathbf{Y}_b$; $\{\mathbf{Y}_b$ is the submatrix of $\mathbf{Y}$ corresponding to the block, c is a preset number.$\}$

5: **While** stopping criterion has not been met **do**

6: Find the indices of c members that best approximate all the residuals: $\mathbf{S}_b' = \{$c indices corresponding to the c largest numbers in $\|(\mathbf{R}^{k-1})^T \mathbf{A}_i\|_2, i = 1, \cdots, m\}$

7: Set $\mathbf{S}_b = \mathbf{S}_b \bigcup \mathbf{S}_b'$

8: Compute $\mathbf{P} = (\mathbf{A}_{\mathbf{S}_b}^T \mathbf{A}_{\mathbf{S}_b})^{-1} \mathbf{A}_{\mathbf{S}_b}^T \mathbf{Y}_b$; $\{\mathbf{A}_{\mathbf{S}_b}$ is the matrix containing the columns of $\mathbf{A}$ having the indices from $\mathbf{S}_b\}$

9: Update the index set $\mathbf{S}_b = \{c$ indices corresponding to the c largest numbers in $\|\mathbf{p}_i^T\|_2, i = 1, \cdots, 2c\}$; $\{\mathbf{p}_i$ is the ith row of $\mathbf{P}\}$

10: Calculate the residual: $\mathbf{R}^k = \mathbf{Y}_b - \mathbf{A}_{\mathbf{S}_b}(\mathbf{A}_{\mathbf{S}_b}^T \mathbf{A}_{\mathbf{S}_b})^{-1} \mathbf{A}_{\mathbf{S}_b}^T \mathbf{Y}_b$

11: $k \leftarrow k + 1$

12: **End while**

13: Set $\mathbf{S} = \mathbf{S} \bigcup \mathbf{S}_b$

14: **End for**

15: Estimate abundances using the original hyperspectral data matrix and spectral library matrix:
$\mathbf{X} \leftarrow \arg\min_{\mathbf{X}} \|\mathbf{A}_{\mathbf{S}}\mathbf{X} - \mathbf{Y}\|$, subject to $\mathbf{X} \geq 0$

matching pursuit (SMP) is summarized in Algorithm 3.

The algorithm includes two main parts: endmember selection and abundance estimation. In the first part, the normalized hyperspectral data and normalized spectral library will be used while the original ones are used for estimating the abundances. In step 8, SMP finds the signature in the spectral library that is best correlated to the residual of the mixed pixel. If the correlation is high enough, in other words, $\mathbf{d}_c$ is greater than a threshold t in step 10, SMP adds the index of the signature into the support $\mathbf{S}$. Step 11 and step 12 ensure that at least one potential endmember is added to $\mathbf{S}$ in each main iteration. So, after one main iteration, several potential endmembers are selected and they span a subspace. Then the hyperspectral data are reconstructed using the selected signatures and the errors of reconstruction are computed (step 14). The iteration stops when convergence is achieved.

We give several possible stopping conditions for SMP:

- The Frobenius norm of the residual $\mathbf{R}^i$ achieves a zero-value or a value below a given threshold δ, for example,

$$\|\mathbf{R}^i\|_F \leq \delta \tag{13}$$

Algorithm 3 Subspace matching pursuit for hyperspectral sparse unmixing

Preprocessing:

Apply zero-mean normalization and l_2 unit-length normalization to the columns of hyperspectral data matrix $\mathbf{Y}$ and spectral library matrix $\mathbf{A}$

Initialization:

Iteration: $i = 0$

Initial residual: $\mathbf{R}^0 = \mathbf{Y}$

Initial support of selected indices: $\mathbf{S} = \phi$

Main iteration (endmember selection):

Update iteration: $i \leftarrow i + 1$

Initialize the index of pixel: $c = 0$

Subiteration:

Update the index of pixel: $c \leftarrow c + 1$

$\mathbf{h}_c \leftarrow \arg\max_{1 \le p \le m} |\mathbf{A}_p^T \mathbf{R}_c^{i-1}|$; {$\mathbf{A}_p$ is the pth column of $\mathbf{A}$, $\mathbf{R}_c^{i-1}$ is the cth column of $\mathbf{R}^{i-1}$, $\mathbf{h}_c$ stores the index of the member in the spectral library that is best correlated to $\mathbf{R}_c^{i-1}$, $\mathbf{h} \in R^K$.}

$\mathbf{d}_c \leftarrow |\mathbf{A}_{\mathbf{h}_c}^T \mathbf{R}_c^{i-1}|$; {$\mathbf{d}_c$ stores the absolute correlation between the current residual for the cth pixel and its best correlated signature, $\mathbf{d} \in R^K$}

If $\mathbf{d}_c \ge t$, update support: $\mathbf{S} \leftarrow \mathbf{S} \cup \{\mathbf{h}_c\}$; {$t$ is a preset threshold}

Stop subiteration if $c = K$ is satified

$idx \leftarrow \arg\max_{1 \le c \le K} \mathbf{d}_c$

Update support: $\mathbf{S} \leftarrow \mathbf{S} \cup \{\mathbf{h}_{idx}\}$

Compute $\mathbf{P} = (\mathbf{A}_\mathbf{S}^T \mathbf{A}_\mathbf{S})^{-1} \mathbf{A}_\mathbf{S}^T \mathbf{Y}$; {$\mathbf{A}_\mathbf{S}$ is the matrix containing the columns of $\mathbf{A}$ having the indices from $\mathbf{S}$}

Calculate the residual: $\mathbf{R}^i = \mathbf{Y} - \mathbf{A}_\mathbf{S}\mathbf{P}$

Stop main iteration if the stopping condition is reached

Inversion (abundance estimation)

Compute the abundance fractions for the scene using the original hyperspectral data and spectral library: $\mathbf{X} \leftarrow \arg\min_\mathbf{X} \|\mathbf{A}_\mathbf{S}\mathbf{X} - \mathbf{Y}\|$, subject to $\mathbf{X} \ge 0$.

- There is little or no improvement between successive main iterations in the minimization of the Frobenius norm of the residual, for example,

$$|\|\mathbf{R}^i\|_F - \|\mathbf{R}^{i-1}\|_F| \le \delta \tag{14}$$

or

$$\frac{|\|\mathbf{R}^i\|_F - \|\mathbf{R}^{i-1}\|_F|}{\|\mathbf{R}^{i-1}\|_F} \le \delta \tag{15}$$

- The number of iterations achieves or exceeds a predefined maximum number of iterations.

In practice, the iterations usually continue until some combinations of stopping conditions are satisfied.

The proposed algorithm is inspired by SOMP and SSP, nevertheless they are quite different. Firstly, SOMP and SSP attempt to find one member or a fixed number of members to maximize the sum of absolute correlations with all the input signals in each iteration, but SMP selects at most one member for each input signal separately. If the correlation of the member SMP selects with the corresponding mixed pixel is high enough, the corresponding index is added to the support. Secondly, SOMP selects one member in each iteration to reconstruct the hyperspectral image while SMP selects at least one member in each iteration. Thirdly, the size of the index support in SSP is fixed but the one in SMP increases after each iteration.

The SMP algorithm is also different from the CoSaMP algorithm [47, 48]. Even though both algorithms select multiple endmembers in every step and share the same residual building process, SMP uses a threshold to decide whether to select one member from the spectral library while CoSaMP selects a fixed number of members in each iteration. Besides, CoSaMP will drop the smallest entries in the least-squares signal approximation while SMP will not. Moreover, SMP utilizes the strong correlation of the pixels in the hyperspectral image yet CoSaMP focuses on one input signal. Thus, as OMP, the performance of CoSaMP will be limited by the high correlation of the spectral library.

There are two main advantages to select more than one endmember in each iteration. On one hand, a faster convergence rate can be obtained. On the other hand, in each iteration, SMP looks for some low-degree mixed pixels from the residual (i.e. $\mathbf{R}^i = \mathbf{Y} - \mathbf{A_S P}$) and extracts several corresponding endmembers. If each endmember has a pure pixel at least, SMP can extract all the endmembers only by one iteration. However, if only one endmember is selected in each iteration, the structures of the other low-degree mixed pixels (or even pure pixels) which are not related to the selected endmember could be destroyed in the reconstruction step. The low-degree mixed pixel here denotes a pixel that has a dominant endmember. For example, if the abundance of an endmember in a mixed pixel is large enough (such as more than 0.6 or 0.7), we can call this endmember a dominant endmember and call the mixed pixel a low-degree mixed pixel.

B. Dictionary Pruning Using SMP

Now we present the SMP as a dictionary pruning algorithm. Note that in the final step of Algorithm 3 a constrained least squares technique is applied to obtain the abundances of the selected potential endmembers. However, by doing so, we do not get the optimal sparse solution. In other words, it is a nonnegative constrained least squares solution using a relatively small library. This line of analysis leads us to consider using the SMP as a dictionary pruning algorithm. Specifically, after obtaining the final support of selected endmembers $\mathbf{S}$, we can use the sub-library $\mathbf{A_S}$ as input for the other algorithms such as SUnSAL-TV and BI-ICE. In this way, we can expect that the SMP algorithm is able to boost the performances of other methods, making them more time-efficient and accurate.

C. Theoretical Analysis

In this section, we will analyse the conditions under which SMP will recover all the endmembers present in the hyperspectral image from the spectral library. Firstly, the knowledge of matrix (or operator) norm is reviewed and

some notations are defined. Then we will analyze the optimal endmember recovery conditions in the noiseless case and noisy case respectively.

Before providing the theorems, we introduce some necessary notations that will be used in the theorems.

Definition 4.1: The l_p norm of a vector $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p} \quad \text{for} \quad 1 \leq p < \infty \tag{16}$$

and

$$\|\mathbf{x}\|_\infty = \max_i |x_i| \tag{17}$$

where x_i is the ith element of $\mathbf{x}$.

Definition 4.2: The (p, q) matrix (or operator) norm of $\mathbf{A}$ is defined as

$$\|\mathbf{A}\|_{p,q} = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_q}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_q \tag{18}$$

Several of the (p, q) matrix norms can be computed easily using the following lemma [43, 44].

Lemma 4.1: Consider matrix $\mathbf{A}$

1) The $(1, q)$ matrix norm is the maximum l_q norm of any column of $\mathbf{A}$.
2) The $(2, 2)$ matrix norm is the maximum singular value of $\mathbf{A}$.
3) The (p, ∞) matrix norm is the maximum $l_{p'}$ norm of any row of $\mathbf{A}$, where $1/p + 1/p' = 1$.

Thus, we have

$$\|\mathbf{A}\|_{1,\infty} = \max_{i,j} |a_{ij}| \tag{19}$$

where a_{ij} is the element of $\mathbf{A}$ that has coordinate (i, j).

Two parameters are often used to describe the library $\mathbf{A}$, i.e. coherence [49] and cumulative coherence [33, 50]. The most fundamental quantity associated with a library is the coherence μ.

Definition 4.3: The coherence μ of a spectral library $\mathbf{A}$ equals the maximum absolute inner product between two distinct spectral signatures

$$\mu = \max_{j\neq k} |\langle \mathbf{A}_k, \mathbf{A}_j \rangle| \tag{20}$$

where $\mathbf{A}_k$ is the kth column of $\mathbf{A}$.

The coherence parameter does not characterize a library very well because it only reflects the most extreme correlations between the signatures in the library. By contrast, the cumulative coherence, which measures the maximum total coherence between a fixed signature and a collection of other signatures, can overcome the drawback of coherence.

Definition 4.4: For a positive integer P, the cumulative coherence is defined as

$$\mu_1(P) = \max_{|\mathbf{\Omega}|=P, \mathbf{\Omega}\subset\mathbf{S}_{all}} \max_{j\notin\mathbf{\Omega}} \sum_{i\in\mathbf{\Omega}} |\langle \mathbf{A}_j, \mathbf{A}_i \rangle| \tag{21}$$

where $\mathbf{S}_{all}$ is the support of all the indices in the library. Specially, $\mu_1(0) = 0$.

From a signal processing viewpoint, here we put forward conditions under which the subspace matching pursuit (SMP) algorithm will recover endmembers present in the hyperspectral data from the spectral library. In the following analysis, assume each column of the hyperspectral data matrix $\mathbf{Y}$ and the spectral library matrix $\mathbf{A}$ has l_2-norm unit length. We first consider the ideal noiseless case.

Definition 4.5: Given a hyperspectral data matrix $\mathbf{Y} \in R^{L\times K}$, a spectral library matrix $\mathbf{A} \in R^{L\times m}$,

- $\mathbf{S}_{opt}$ is a subset of $\mathbf{S}_{all}$ with cardinality P and it lists all the P endmembers present in the hyperspectral scene;
- $\mathbf{A}_{opt}$ is a submatrix of the library matrix $\mathbf{A}$ and the columns of $\mathbf{A}_{opt}$ are the P endmembers present in the hyperspectral image, $\mathbf{A}_{opt} \in R^{L\times P}$;
- $\mathbf{A}_{err}$ is the matrix whose columns are the $(m-P)$ remaining signatures in the spectral library that are not present in the hyperspectral image, $\mathbf{A}_{err} \in R^{L\times(m-P)}$;
- $\mathbf{Y}_{rec}$ is the reconstruction of $\mathbf{Y}$ by SMP after some iterations and $\mathbf{Y}_{rec}$ also lies in the column range of $\mathbf{A}_{opt}$ (assume all the previous signatures SMP selects are optimal endmembers);
- $\mathbf{R}$ is the residual of $\mathbf{Y}$ after reconstruction, $\mathbf{R} = \mathbf{Y} - \mathbf{Y}_{rec}$.

With the definitions above in hand, we give out the recovery theorem in the noiseless case.

Theorem 4.1: In the noiseless case, a sufficient condition for SMP to recover all the endmembers listed in the support $\mathbf{S}_{opt}$ is that

$$\max_{j\notin\mathbf{S}_{opt}} \|\mathbf{A}_{opt}^{\dagger}\mathbf{A}_j\|_1 < t \tag{22}$$

where $\mathbf{A}_j$ is the jth column of $\mathbf{A}$, $t \leq 1$ is the threshold at step 10 in the SMP algorithm and $\mathbf{A}_{opt}^{\dagger}$ is the pseudoinverse of matrix $\mathbf{A}_{opt}$. It is defined as $\mathbf{A}_{opt}^{\dagger} = (\mathbf{A}_{opt}^T\mathbf{A}_{opt})^{-1}\mathbf{A}_{opt}^T$.

The proof of this theorem is shown in Appendix A. If Theorem 4.1 holds, SMP will identify all the P endmembers that make up the optimal representation of $\mathbf{Y}$ in at most P iterations.

In the practical applications, the hyperspectral data are always affected by noise. In the noisy case, assume $\mathbf{Y}_{opt} \in R^{L\times K}$ is an optimal P-term approximation of $\mathbf{Y}$ and it lies in the column range of $\mathbf{A}_{opt}$. Then we can develop a condition under which SMP will recover the optimal endmembers from the spectral library in the noisy case.

Theorem 4.2: Suppose SMP has picked k optimal endmembers from the spectral library and $\mu_1(P) < \frac{t}{1+t}$, $M(k) < \frac{t}{1+t}$, then a sufficient condition for SMP to recover only optimal endmembers listed in the support $\mathbf{S}_{opt}$ in the next iteration is that

$$(t - \frac{M(k)}{1-M(k)})\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} > \|\mathbf{A}_{err}^T(\mathbf{Y} - \mathbf{Y}_{opt})\|_{1,\infty} \tag{23}$$

where $M(k) = M(k,P) = \frac{\mu_1(P-k)}{1-\mu_1(k)}$.

The proof of Theorem 4.2 is shown in Appendix B. This theorem suggests that the SMP algorithm will behave better when the number of endmembers that construct the hyperspectral data is not too large and the spectral library is less correlated. Thus the block-processing strategy is applicable for the SMP algorithm. Note that in some cases, the prerequisite $\mu_1(P) < \frac{t}{1+t}$ of the spectral library might not be satisfied, especially when the P gets large.

However, i) we can adopt the block-processing strategy in Section III to make the number of endmembers present in the block smaller; ii) the experimental results indicate that the theorem is very conservative; iii) it can be observed in the available spectral libraries that the spectral signatures appear in the form of groups (e.g. alterations of a single spectral signature in the USGS spectral library) [51] and the correlation between different groups is weaker than that within all the signatures in a spectral library. Thus, if we only need to estimate the abundances for the different groups rather than all the signatures in the spectral library, we can expect SMP to get a better performance.

V. Experiments

In this section, we explore the potential of the SGA methods for sparse unmixing. Three synthetic data sets and one real-world data set are used to test the performances of the different algorithms. All the considered algorithms have taken into account the abundance nonnegativity constraint. The preprocessing procure adopted by the SGA methods is also applied to the OMP algorithm and the CoSaMP algorithm to improve their performances. The stopping conditions we choose for SMP and SOMP in the experiments are the combination of the second condition (inequality (15)) and the third condition described in Section IV. The stopping conditions for SSP are a little different: it terminates when the residual begins to increase, or the number of iterations achieves or exceeds a predefined maximum number of iterations. The constrained least squares problem in the final step of the three SGA methods can be solved using the active set method. Specifically, we use the *lsqnonneg* function in MATLAB to solve it. The TV regularizer used in SUnSAL-TV is a non-isotropic one. All the parameters of the considered algorithms are tuned to the best performances. Complexity analysis and discussion of the parameters setting come after the subsection of real data experiment.

The root mean square error (RMSE) [52] is used to evaluate the abundance estimations. For the ith endmember, RMSE is defined as

$$\text{RMSE}_i = \sqrt{\frac{1}{K}\sum_{j=1}^{K}(\mathbf{X}_{ij} - \hat{\mathbf{X}}_{ij})^2}. \tag{24}$$

Here, $\mathbf{X}$ denotes the true abundances and $\hat{\mathbf{X}}$ represents the estimated ones. 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.

A. Evaluation with Synthetic Data 1

For our first two synthetic data experiments, we evaluate the performances of the sparse unmixing algorithms in situations of different noise types, different signal-to-noise ratios (SNR $\equiv 10\log_{10}\frac{\|\mathbf{Ax}\|_2^2}{\|\mathbf{n}\|_2^2}$) of noise and different endmember numbers. The first synthetic hyperspectral image is homogeneous with rich spatial-contextual information but the second one is randomly generated. The third synthetic data experiment is designed to see whether the three SGA methods will miss the actual endmembers which are not very prominent in any of the blocks. The spectral library used in all the synthetic experiments is the United States Geological Survey (USGS) [53] digital spectral library. The reflectance values of 498 materials are measured for 224 spectral bands distributed uniformly

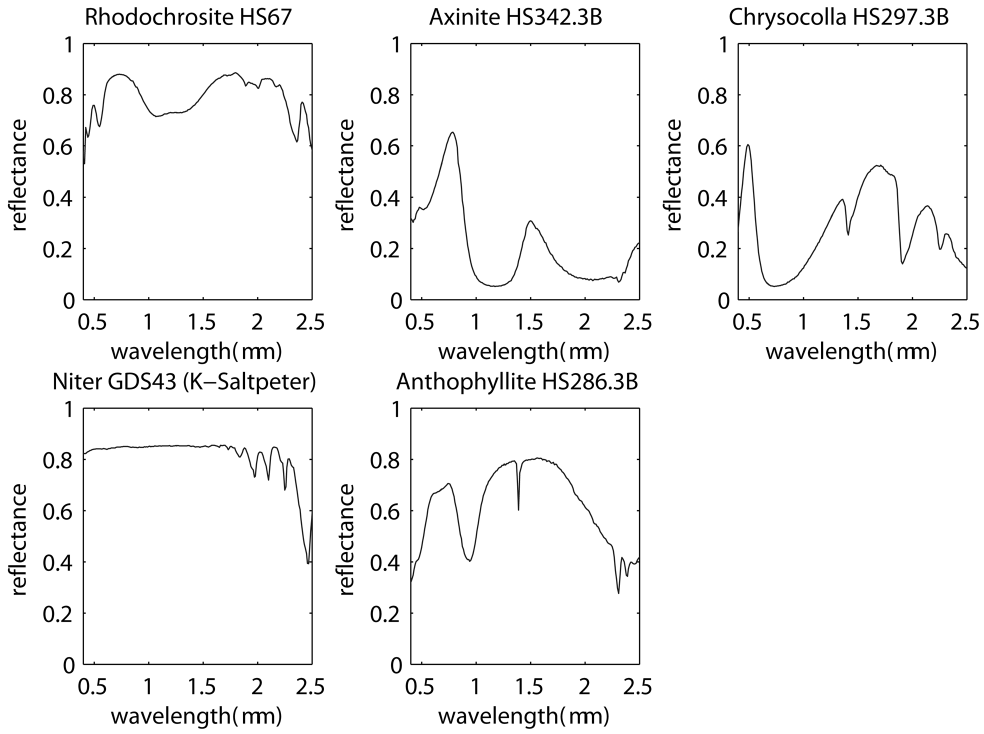

Fig. 1. Five example spectral signatures from USGS used in our synthetic data experiments. The title of each subimage denotes the mineral corresponding to the signature.

in the interval $0.4-2.5\ \mu m$ ($\mathbf{A} \in R^{224\times 498}$). Fifteen spectral signatures are chosen from the library to generate our synthetic data. Fig. 1 shows five example endmember signatures used for all the following experiments. The other ten endmembers that are not displayed in the figure include Neodymium Oxide GDS34, Monazite HS255.3B, Samarium Oxide GDS36, Pigeonite HS199.3B, Meionite WS700.HLsep, Spodumene HS210.3B, Labradorite HS17.3B, Grossular WS484, Zoisite HS347.3B and Wollastonite HS348.3B, which are used to evaluate the performances of the sparse unmixing algorithms in situations of different endmember numbers.

The first 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 P. The size of spectral signatures matrix $\mathbf{W}$ is $L \times P$ ($L = 224$).
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 P endmembers in the scene and specific values are stored in $\mathbf{H}$ with a size of of $P \times K$ ($K = z^2 \times z^2$).

4) Use linear spectral mixing model $\mathbf{Y} = \mathbf{W} \times \mathbf{H}$ to generate hyperspectral data, add Gaussian white noise or correlated noise[3] with specific SNR at the same time. The size of hyperspectral data $\mathbf{Y}$ is $L \times K$.

Note that there is no pure pixel in this synthetic hyperspectral data set. We use this homogeneous data set because there is always spatial-contextual information within the real hyperspectral image. This situation is beneficial to the SGA methods and SUnSAL-TV because they can exploit the spatial information of the hyperspectral data. SOMP, SSP and SMP are compared with OMP, SUnSAL, CSUnSAL, SUnSAL-TV and CoSaMP. Besides, we also use SMP as a dictionary pruning algorithm to improve the performance of SUnSAL-TV (denoted by SMP-SUnSAL-TV).

We first analyse without considering the SMP-SUnSAL-TV. Fig. 2 shows results obtained on *Synthetic Data 1* with white noise as function of SNR. The endmember number is 5. We can find that SMP, SOMP, SSP and SUnSAL-TV behave better than OMP, CoSaMP, SUnSAL and CSUnSAL. This result indicates that for this homogeneous data set the combination of spatial information into the sparse unmixing algorithms can improve their performances. The CoSaMP algorithm does not behave well because the correlation of the spectral library is very large (close to 1), which violates the assumption of the CoSaMP algorithm that the library matrix has restricted isometry constant far less than 1 [47]. All the three SGA methods have comparable performances with SUnSAL-TV, which gives the evidence of their potential. SMP can always obtain the best estimation of the abundances. Besides, during this experiment, to pick up all the actual endmembers, SSP and SOMP need to adopt the block-processing strategy but SMP does not need. Fig. 3 shows results obtained on this data set with white noise as function of endmember number. The SNR of the noise is 30 dB. As shown, most algorithms tend to get worse performances as the number of endmembers gets larger. However, we can still find that the performances of the SGA methods are comparable with that of SUnSAL-TV, and SMP behaves best when the number of endmembers is less than 15. In this experiment, when the number of endmembers is larger than 3, SSP and SOMP need to adopt the block-processing strategy to avoid missing the actual endmembers; SMP needs to adopt the strategy when the number of endmembers is larger than 7. This observation indicates that the prediction made by Theorem 4.2 is rather conservative as $\mu_1(7)$ of the spectral library we use is much larger than $\frac{1}{2}$.

Fig. 4 and Fig. 5 show the results obtained on the *Synthetic Data 1* affected by correlated noise as functions of SNR and endmember number, respectively. As the noise in the hyperspectral images is usually correlated, this case is closer to the practical ones. As shown in Fig. 4, SOMP gets an overall better performance in the correlated noise case than in the white noise one. The performances of SMP, SSP, SUnSAL, OMP and CoSaMP in both cases are nearly the same. However, SUnSAL-TV and CSUnSAL behave worse when the data are corrupted by correlated noise. The results shown in Fig. 5 are similar with those in white noise case.

Fig. 6 shows the number of potential endmembers retained from the original library by the SGA methods. In all the cases, the SGA methods succeed to find all the actual endmembers. It can be observed that SMP can select the

[3]The Gaussian white noise is generated using the *awgn* function in MATLAB. The correlated noise is generated using the *correlatedGaussianNoise* function that is available online: http://www.mathworks.com/matlabcentral/fileexchange/21156-correlated-gaussian-noise/content/correlatedGaussianNoise.m. The correlation matrix is set as default.

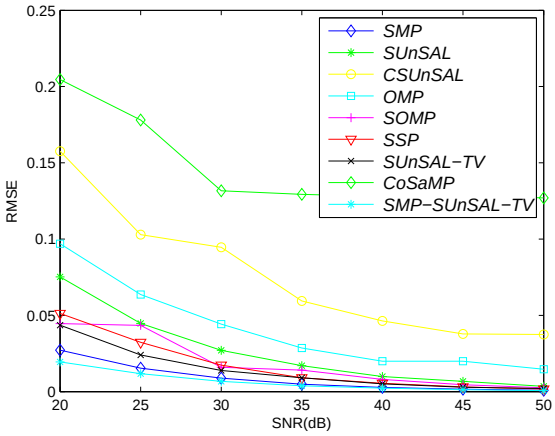

Fig. 2. Results on *Synthetic Data 1* with white noise when the endmember number is 5: RMSE as function of SNR.

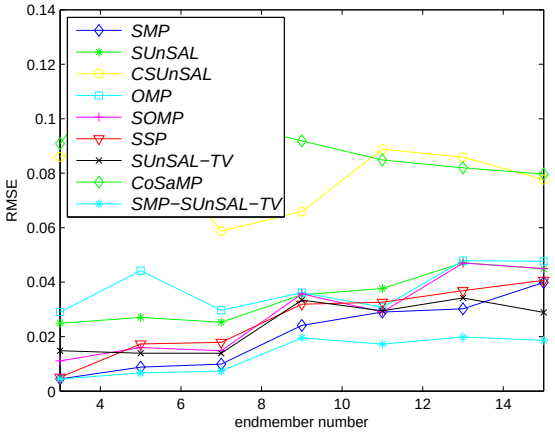

Fig. 3. Results on *Synthetic Data 1* with 30 dB white noise: RMSE as function of endmember number.

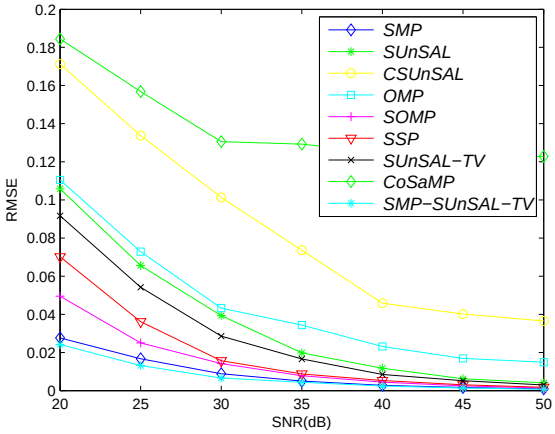

Fig. 4. Results on *Synthetic Data 1* with correlated noise when the endmember number is 5: RMSE as function of SNR.

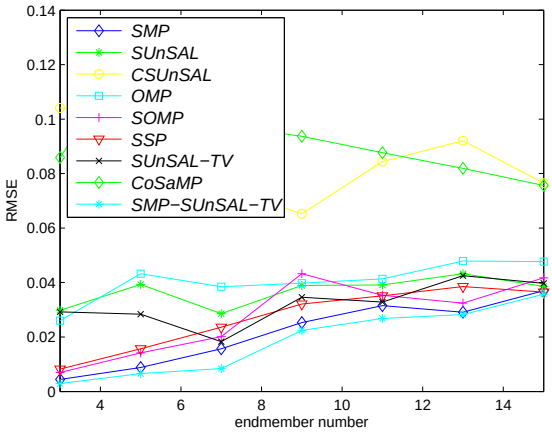

Fig. 5. Results on *Synthetic Data 1* with 30 dB correlated noise: RMSE as function of endmember number.

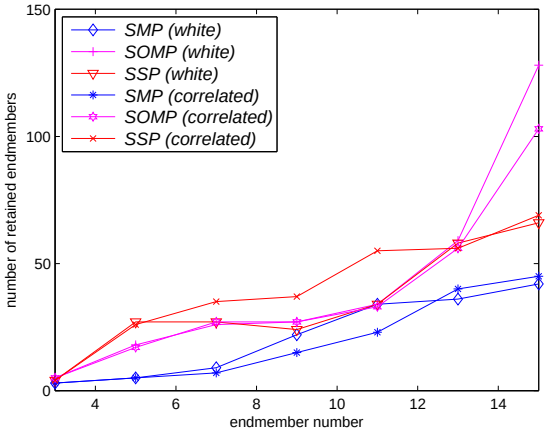

Fig. 6. Results obtained by the three SGA methods on *Synthetic Data 1* with 30 dB white noise or correlated noise: the number of retained endmembers as function of endmember number.

actual endmembers more accurately than SOMP and SSP, which is also the reason why SMP behaves better than SOMP and SSP.

Now, we turn to the SMP-SUnSAL-TV. From Figs. 2-5 we can observe two facts: i) when serving as a dictionary pruning algorithm, SMP can improve the performance of SUnSAL-TV; ii) the SMP-SUnSAL-TV behaves better than SMP, which suggests that combination of SMP with the other sparse unmixing algorithms can be better than both the two algorithms. Fig. 7 shows the true abundance maps and the abundance maps estimated by the different algorithms on *Synthetic Data 1* with 30 dB white noise when the endmember number is 5.

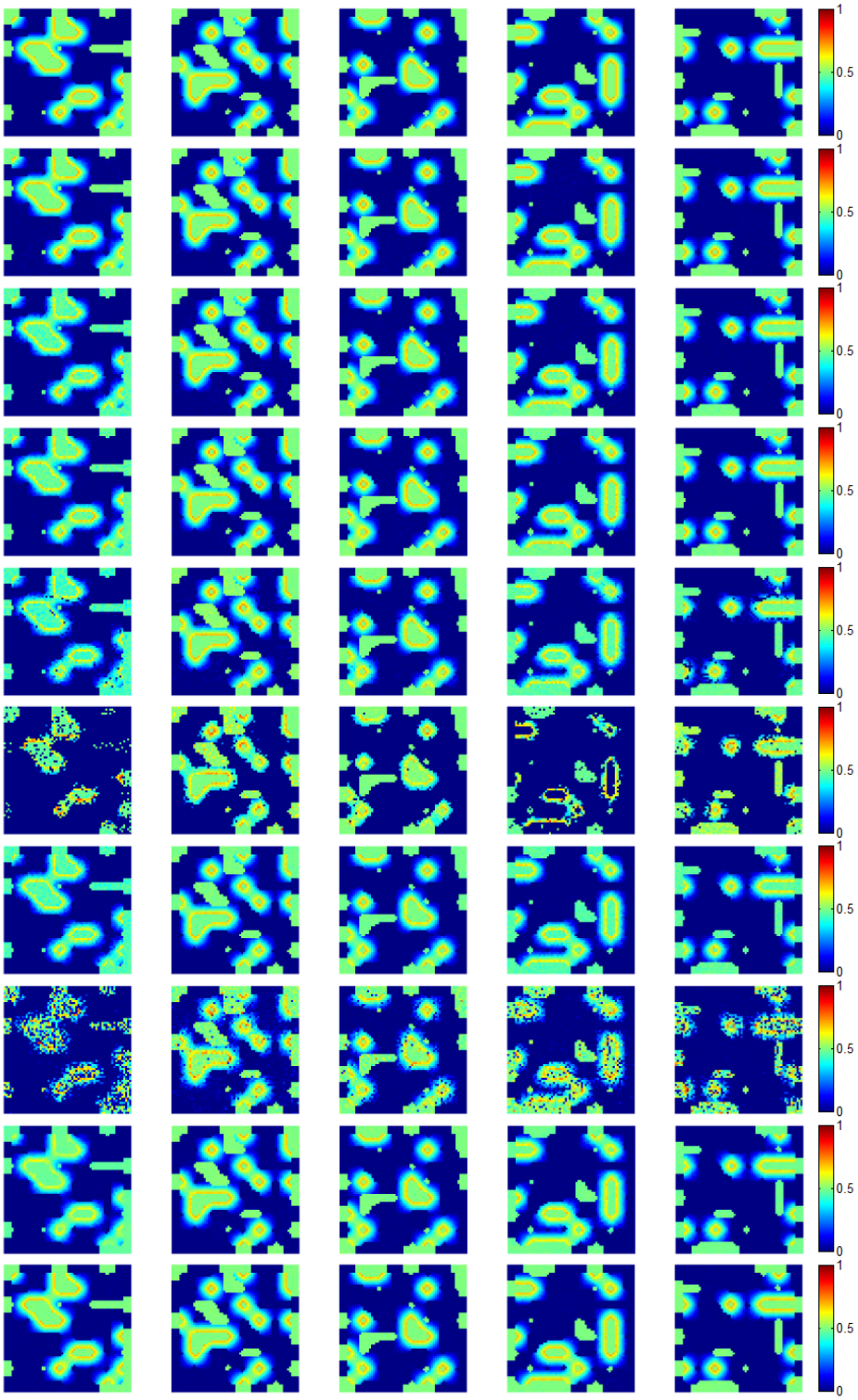

Fig. 7. Comparison of abundance maps on *Synthetic Data 1* with 30 dB white noise when the endmember number is 5. From top row to bottom row are true abundance maps, abundance maps obtained by SMP, abundance maps obtained by SOMP, abundance maps obtained by SSP, abundance maps obtained by OMP, abundance maps obtained by CoSaMP, abundance maps obtained by SUnSAL, abundance maps obtained by CSUnSAL, abundance maps obtained by SUnSAL-TV, and abundance maps obtained by SMP-SUnSAL-TV, respectively.

B. Evaluation with Synthetic Data 2

The second synthetic data set is randomly generated following a Dirichlet distribution [9, 25, 26]. However, as there exist pure pixels for each endmember in this data set, it becomes extremely easy for SMP to pick up all the actual endmembers in only one iteration. Thus, similar as the *Synthetic Data 1*, the abundances are forced to be smaller than 0.7. The size of this data set is $30 \times 30 \times 224$. The BI-ICE algorithm is included for comparison. The SUnSAL-TV algorithm is excluded as this data set does not exhibit any spatial coherence. As the pixels in a neighborhood are hardly correlated, it is much more difficult for the SGA methods to unmix compared with the first data set. Similarly, we use SMP as a dictionary pruning algorithm to improve the performances of SUnSAL and BI-ICE (denoted by SMP-SUnSAL and SMP-BI-ICE, respectively). The curves corresponding to CoSaMP is removed to make the comparison of the other algorithms clearer.

We first analyse without considering the SMP-SUnSAL and SMP-BI-ICE. Fig. 8 and Fig. 9 show the results obtained on the *Synthetic Data 2* affected by white noise as functions of SNR and endmember number, respectively. The performances of the SSP and SOMP on this inhomogeneous data set degrade compared with those on the homogeneous one. Compared with SOMP and SSP that rely on the overall contribution of the actual endmembers to the block, SMP utilizes the low-degree mixed pixels in the block to reconstruct the block step by step, thus SMP still behaves well on this inhomogeneous data set. When the endmember number is smaller than 9, the BI-ICE algorithm has similar performance with SMP. As the endmember number increases, SMP behaves better than BI-ICE.

Fig. 10 and Fig. 11 show the results obtained on the *Synthetic Data 2* affected by correlated noise as functions of SNR and endmember number, respectively. Some similar observations can be made as in the *Synthetic Data 1*. Here, we draw attention to the performance of BI-ICE algorithm. It is obvious that BI-ICE algorithm behaves much better in the white noise case than in the correlated noise case. This is because in the Bayesian model that the BI-ICE algorithm solves, the prior assumed for the noise is a zero-mean Gaussian distributed random vector, with independent and identically distributed (i.i.d.) elements. As the assumption no longer holds in the correlated noise case, the performance of BI-ICE degrades.

Fig. 12 shows the number of potential endmembers retained from the original library by the SGA methods. Different with Fig. 6, when the endmember number is 15, SMP misses one actual endmember in the correlated noise case; SOMP misses one actual endmember in both white noise case and correlated noise case. Nevertheless, the results shown in Fig. 12 correspond to the best RMSEs. If the block sizes of SMP and SOMP are set smaller, they will select all the actual endmembers, but the RMSEs will increase. This phenomenon reminds us that the block sizes of the SGA methods should be set carefully, which will be discussed later. Generally speaking, SMP can select the actual endmembers more accurately than SOMP and SSP.

Now consider the SMP-SUnSAL and SMP-BI-ICE. The block size of the SMP is set smaller to guarantee that all the actual endmembers will be retained. From Figs. 8-11, we can find that the performances of the SUnSAL and BI-ICE improve when using the SMP as a dictionary pruning algorithm. Besides, among all the algorithms,

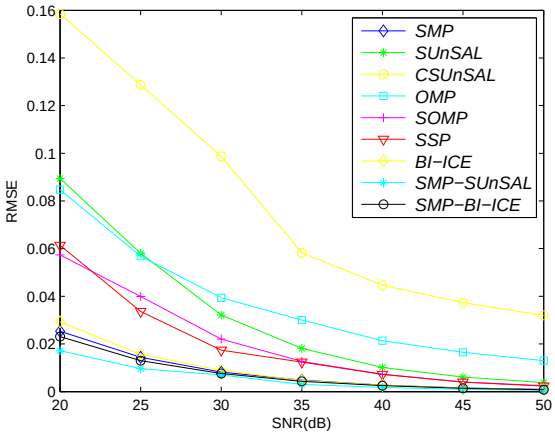

Fig. 8. Results on *Synthetic Data 2* with white noise when the endmember number is 5: RMSE as function of SNR.

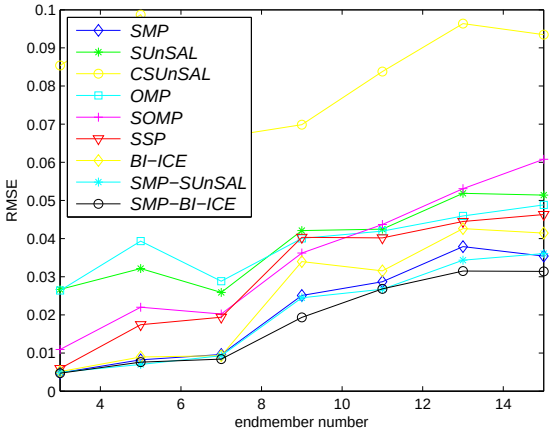

Fig. 9. Results on *Synthetic Data 2* with 30 dB white noise: RMSE as function of endmember number.

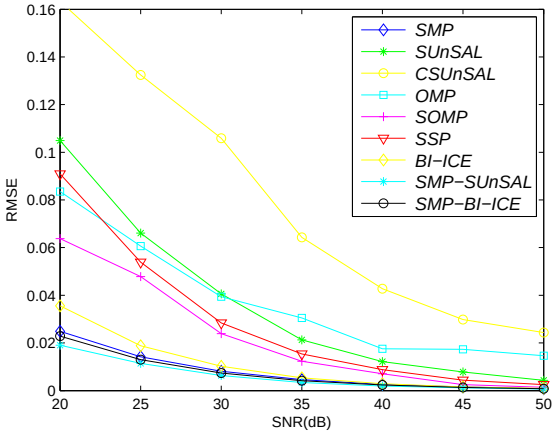

Fig. 10. Results on *Synthetic Data 2* with correlated noise when the endmember number is 5: RMSE as function of SNR.

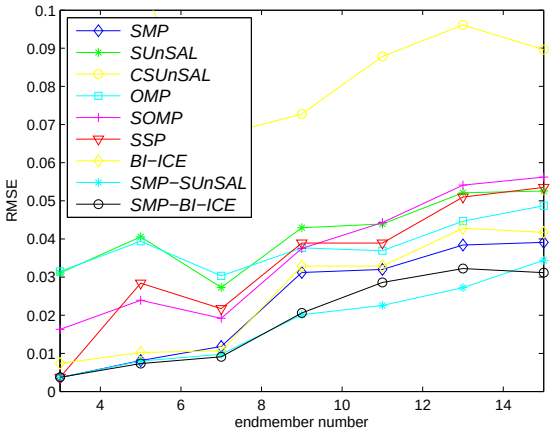

Fig. 11. Results on *Synthetic Data 2* with 30 dB correlated noise: RMSE as function of endmember number.

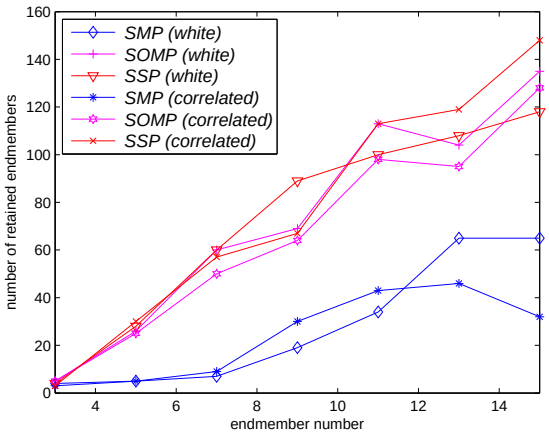

Fig. 12. Results obtained by the three SGA methods on *Synthetic Data 2* with 30 dB white noise or correlated noise: the number of retained endmembers as function of endmember number.

the best results are always obtained by SMP-SUnSAL or SMP-BI-ICE.

C. Evaluation with Synthetic Data 3

In this subsection, a toy synthetic hyperspectral data set is generated to test whether the SGA methods will miss the actual endmembers which are not very prominent in any of the blocks. In such circumstance, the experimental results are prone to be affected by the noise due to the very small abundance fractions of these endmembers. Thus, the experiment is designed as follows. A $10 \times 10 \times 224$ toy hyperspectral data set following the Dirichlet distribution is generated using five randomly selected endmembers. We choose the size 10×10 so that the generated hyperspectral data can simulate a block. The abundance fractions of one or two endmembers are forced to be smaller than 0.2 or 0.1 and the abundance fractions of the other four or three endmembers are scaled to satisfy the sum-to-

TABLE I

THE PROBABILITIES THAT DIFFERENT ALGORITHMS WITH DIFFERENT BLOCK SIZES SELECT ALL THE ACTUAL ENDMEMBERS WHEN THE ABUNDANCE FRACTIONS OF ONE OR TWO ENDMEMBERS ARE BELOW 0.2 OR 0.1.

		One endmember		Two endmembers	
Algorithm	Block size	Fraction below 0.2	Fraction below 0.1	Fractions below 0.2	Fractions below 0.1
SMP	10	0.8	0.7	0.7	0.5
	5	1	1	1	0.8
	3	1	1	1	0.9
SOMP	10	0.4	0.3	0.3	0.1
	5	0.7	0.6	0.6	0.5
	3	0.9	0.8	0.9	0.8
SSP	10	0.1	0.1	0.1	0
	5	0.2	0.2	0.2	0
	3	0.5	0.4	0.4	0.3

one constraint. This toy hyperspectral data set is corrupted by 30 dB white noise. Then the three SGA methods are tested on this data set using different block sizes (i.e. 10, 5 and 3). The probabilities that these algorithms select all the actual endmembers are computed after 10 runs. From Tab. I, we can find that it will be more difficult for the SGA methods to select all the actual endmembers when the abundance fractions of the endmembers that are not prominent get smaller and the number of such endmembers becomes larger. Note that in this data set, the convex relaxation methods also can not estimate the abundances of the endmembers that are not prominent well. However, for the SGA methods, if the block size is set smaller, the probability of selecting all the actual endmembers will increase. Generally speaking, SMP behaves much better than SOMP and SSP in this experiment.

D. Evaluation with Real Data

Now, we test the validity of the sparse unmixing algorithms on the real data. As the true abundance maps do not exist, we resort to the classification maps of the real hyperspectral image to make a qualitative analysis. For visual comparison, we display the abundances estimated by the three SGA methods, SUnSAL-TV and SMP-SUnSAL-TV.

The data set used in this experiment is the well-known AVIRIS Cuprite data set[4]. The hyperspectral scene used in our experiment is a 350×350 pixels subset with 188 spectral bands (low-SNR bands are removed). The spectral library used here is the USGS spectral library (contains 498 minerals) with the corresponding bands removed. The minerals map[5] which was produced by a Tricorder 3.3 software product is shown in Fig. 13. Note that the Tricorder map was produced in 1995, but the publicly available AVIRIS Cuprite data were collected in 1997. Thus, we only adopt the minerals map as a reference to make a qualitative analysis of the performances of different sparse unmixing methods.

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

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

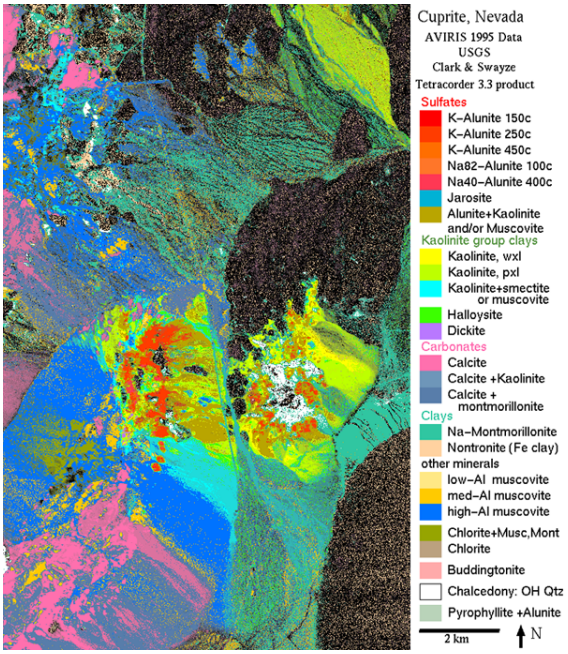

Fig. 13. USGS map showing the distribution of different minerals in the Cuprite mining district in Nevada.

All the three SGA methods have taken the block-processing strategy to get better performances. The block sizes used in SOMP and SSP are both 8. The block size used in SMP is 20. Fig. 14 shows a qualitative comparison among the fractional abundance maps of 3 highly materials in the considered scene estimated by SUnSAL-TV, SMP, SOMP, SSP and SMP-SUnSAL-TV. The distribution maps of these materials produced by Tricorder software are also displayed. The abundance maps estimated by SUnSAL-TV displayed in Fig. 14 are from [29]. 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 image 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 explain why the Tricorder maps and the abundance maps can not be exactly the same. However, we can observe 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 SMP and the other SGA methods are all valid to unmix the real world hyperspectral data.

E. Complexity Analysis and Discussion of Parameters Setting

We first compare the computational complexities of the different sparse unmixing algorithms. In the following analysis, suppose L is the number of bands of the hyperspectral image, K is the number of pixels in the hyperspectral image, m is the number of spectral signatures in the spectral library. For the whole hyperspectral data, SMP, SOMP, SSP, OMP and CoSaMP have complexity $O(mLK)$; SUnSAL, CSUnSAL and SUnSAL-TV have complexity

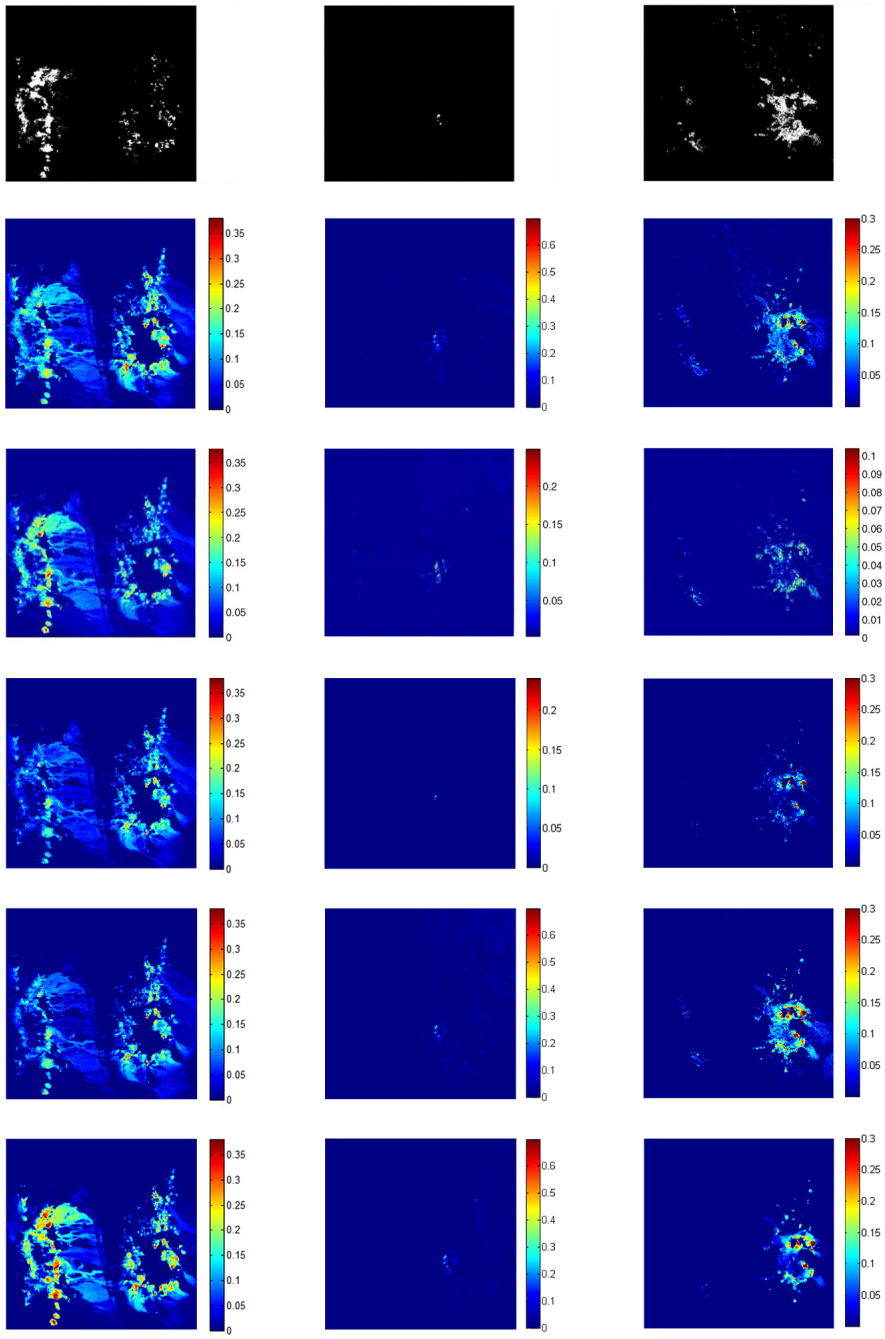

Fig. 14. The fractional abundance maps estimated by SMP, SUnSAL-TV, SOMP, SSP, SMP-SUnSAL-TV and the distribution maps produced by Tricorder software for the 350×350 pixels subset of AVIRIS Cuprite scene. From top row to bottom row are the maps produced or estimated by Tricorder software, SMP, SUnSAL-TV, SOMP, SSP and SMP-SUnSAL-TV respectively. From left column to right column are the maps corresponding to Alunite, Buddingtonite and Chalcedony, respectively.

TABLE II

PROCESSING TIMES MEASURED AFTER APPLYING THE ALGORITHMS TO THE *Synthetic Data 2* WITH 30 DB WHITE NOISE CONSTRUCTED BY 5 ENDMEMBERS

Algorithms	SMP	SOMP	CoSaMP	SUnSAL	SUnSAL-TV	SMP-SUnSAL
Time (s)	1.38	1.36	6.59	9.44	124.46	1.20
Algorithms	SSP	OMP	CSUnSAL	BI-ICE	SMP-BI-ICE	
Time (s)	1.45	6.97	16.73	60114.10	17.70	

$O(L^2K)$; BI-ICE has complexity $O(m^3K)$. As L and m are always the same order of magnitude, BI-ICE is more computational complex than the other algorithms. For example, Tab. II reports the processing time (the average time after running each algorithm 10 times) measured after applying the algorithms to the *Synthetic Data 2* with 30 dB white noise constructed by 5 endmembers. A desktop PC equipped with an Intel Core 2 Duo CPU (2.66 GHz) and 4 GB of RAM memory was used to implement all the algorithms on MATLAB R2010a. The results of the SMP-SUnSAL and SMP-BI-ICE are also included. It can be observed that the BI-ICE algorithm is very time-consuming, which is in accordance with our complexity analysis. This is also the reason why we do not include BI-ICE in the first synthetic experiment. Besides, we can find that taking SMP as a dictionary pruning algorithm, the speeds of SUnSAL and BI-ICE increase dramatically.

Another critical issue concerned with the proposed algorithm is the parameters setting. Two parameters, namely the block size and the threshold t, are important to the SMP algorithm. As mentioned earlier, when the number of endmembers present in the hyperspectral image is large, the SGA methods could miss the actual endmembers. In this paper, we propose to adopt the block-processing strategy with proper block size to sidestep this obstacle. Thus, two questions are to be solved: when the block-processing strategy is necessary and how to properly set the block size. For the first two synthetic data experiments, we have tested the block size from 3 to the image size with the step size of 2. We find that when the number of endmembers present in the block or hyperspectral data is no more than 5, SMP can find the actual endmembers very efficiently. Thus, the block-processing strategy is necessary when the endmember number is larger than 5. But when some endmembers are not prominent in all the pixels of a hyperspectral image, a block-processing strategy will make these endmembers more likely to be selected, which we can infer from the third synthetic experiment. Then a natural answer to the second question is that the block size should be chosen to make the number of endmembers in each block no more than 5. As we only need to select all the actual endmembers from all the blocks rather than to find the actual endmembers present in each block exactly, the choosing of block size can be relaxed to make the average number of endmembers in each block no more than 5. Besides, given a hyperspectral image, smaller block size leads to larger block number, which also helps preventing SMP from missing the actual endmembers that are not prominent in any of the pixels. Furthermore, we can also use overlapping blocks to increase the number of blocks when the block size is fixed. In this way, the probability that SMP misses the endmembers that are not prominent in any of the blocks will decrease further, but the number of endmembers selected by SMP that are not present in the scene will also increase. Several algorithms

have been proposed to estimate the number of endmembers in the hyperspectral image efficiently [54, 55]. In this paper, we use the virtual dimensionality [54] to estimate the number of endmembers in each block in the real data experiment. As the block size decreases, the average number of endmembers in each block will decrease. For the real data used in our experiment, when the block size is 20, the average number of endmembers in all the blocks is 4.5. Thus, the block size we set for SMP in this experiment is 20.

The threshold t in the SMP algorithm is also an important parameter. Smaller t indicates that the algorithm will converge faster and more potential members will be selected in the first few iterations. However, in this situation, some spectral signatures that are not present in the scene could be selected, which would have a negative impact on the results. Thus, t should not be set too small. Theorem 4.2 also suggests that SMP favors large t. Indeed, after testing the performance of SMP using different values of t in different situations, we find that when t is close to 1, SMP behaves well and is insensitive to this parameter. In all the experiments of this paper, t is set fixed to be 0.96, which is enough for SMP to estimate the abundances well. Certainly, a fine-tuning t according to each data set will make SMP behave better.

The block size for SOMP and SSP is much more difficult to set compared with that of SMP. On one hand, the performances of both SOMP and SSP differ a lot in different data sets. On the other hand, when there exist some endmembers in the hyperspectral image that are not prominent in every pixel, SOMP and SSP are more likely to miss them than SMP, which can be inferred from the third synthetic experiment. However, as the real hyperspectral image is always homogeneous, we set the block size for SOMP and SSP in the real data experiment by referring to the results of the first synthetic experiment. As mentioned in Section V-A, SOMP and SSP need to adopt the block-processing strategy when the endmember number is larger than 3. Thus, following the procedure that we set the block size for SMP and considering that the block size for SOMP and SSP should be set smaller to avoid missing the endmembers that are not very prominent in every pixel, finally the block size we set for SOMP and SSP is 8.

VI. Conclusion

In this paper, we present a new greedy algorithm termed subspace matching pursuit (SMP) for sparse unmixing of hyperspectral data. Compared with the existing SGA methods, SMP can make use of the low-degree mixed pixels in the hyperspectral image as well as the spatial information. As a greedy algorithm, SMP also has advantages such as low computational complexity of solving the optimization problems containing non-smooth terms and that it can get an approximate solution for the l_0 problem directly. Furthermore, the proposed algorithm can serve as a dictionary pruning algorithm to boost any other sparse unmixing algorithms, making them more accurate and time-efficient. It is proved that under certain conditions, SMP can recover the optimal endmembers from the spectral library. Experiments on the synthetic data and the real data indicate the potential of using SGA methods for sparse unmixing of hyperspectral data. They also demonstrate the effectiveness of the proposed algorithm.

APPENDIX A
PROOF OF THEOREM 4.1

Firstly, we provide some lemmas.

Lemma A.1: For any matrix $\mathbf{A}$, we have

$$1)\ \|\mathbf{A}\mathbf{x}\|_q \leq \|\mathbf{A}\|_{p,q}\|\mathbf{x}\|_p \tag{25}$$

$$2)\ \|\mathbf{A}^T\|_{\infty,\infty} = \|\mathbf{A}\|_{1,1} \tag{26}$$

Then, we can derive the following property of the (p,∞) matrix norm.

Lemma A.2: The matrices $\mathbf{A}$, $\mathbf{B}$ and scalar $p > 0$ satisfy

$$\|\mathbf{A}\mathbf{B}\|_{p,\infty} \leq \|\mathbf{A}\|_{\infty,\infty}\|\mathbf{B}\|_{p,\infty} \tag{27}$$

Proof: We prove in the similar way as [44]. According to the definition of matrix norm and Lemma A.1, we have

$$\|\mathbf{A}\mathbf{B}\|_{p,\infty} = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{B}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_p} \tag{28}$$

$$\leq \max_{\mathbf{x}\neq 0} \|\mathbf{A}\|_{\infty,\infty} \frac{\|\mathbf{B}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_p} \tag{29}$$

$$= \|\mathbf{A}\|_{\infty,\infty} \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{B}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_p} \tag{30}$$

$$= \|\mathbf{A}\|_{\infty,\infty}\|\mathbf{B}\|_{p,\infty} \tag{31}$$

Thus, the lemma is proved. □

Lemma A.3: Given an invertible matrix $\mathbf{A}$ and an arbitrary matrix $\mathbf{B}$, the following inequality holds for any scalar $p > 0$

$$\|\mathbf{A}\mathbf{B}\|_{p,\infty} \geq \|\mathbf{A}^{-1}\|_{\infty,\infty}^{-1}\|\mathbf{B}\|_{p,\infty} \tag{32}$$

Proof: If $\mathbf{B} = 0$, the inequality holds, or we have

$$\Big(\min_{\mathbf{B}\neq 0} \frac{\|\mathbf{A}\mathbf{B}\|_{p,\infty}}{\|\mathbf{B}\|_{p,\infty}}\Big)^{-1} = \max_{\mathbf{B}\neq 0} \frac{\|\mathbf{B}\|_{p,\infty}}{\|\mathbf{A}\mathbf{B}\|_{p,\infty}} \tag{33}$$

Set $\mathbf{C} = \mathbf{A}\mathbf{B}$. Thus,

$$\max_{\mathbf{B}\neq 0} \frac{\|\mathbf{B}\|_{p,\infty}}{\|\mathbf{A}\mathbf{B}\|_{p,\infty}} = \max_{\mathbf{C}\neq 0} \frac{\|\mathbf{A}^{-1}\mathbf{C}\|_{p,\infty}}{\|\mathbf{C}\|_{p,\infty}} \leq \|\mathbf{A}^{-1}\|_{\infty,\infty} \tag{34}$$

The second inequality can be derived by using Lemma A.2. Combine Eq. (33) and Eq. (34)

$$\Big(\min_{\mathbf{B}\neq 0} \frac{\|\mathbf{A}\mathbf{B}\|_{p,\infty}}{\|\mathbf{B}\|_{p,\infty}}\Big)^{-1} \leq \|\mathbf{A}^{-1}\|_{\infty,\infty} \tag{35}$$

Finally, some algebraic manipulations yield Eq. (32). □

Lemma A.4: Given an input hyperspectral data matrix $\mathbf{Y}$ each column of which has l_2-norm unit length, then during the iteration of the SMP algorithm, the l_2-norm length of each column of the residual matrix $\mathbf{R}$ will be no more than 1.

Proof: Let $\mathbf{y} \in R^L$ be an arbitrary column of $\mathbf{Y}$. Vectors $\mathbf{r}$ and $\mathbf{y}_{rec}$ are the corresponding columns of $\mathbf{R}$ and $\mathbf{Y}_{rec}$ after some iterations, respectively. Then we have the following equality

$$\mathbf{y} = \mathbf{r} + \mathbf{y}_{rec} \tag{36}$$

Note $\mathbf{y}_{rec}$ is the orthogonal projection of $\mathbf{y}$ onto the subspace spanned by the members that have been selected:

$$\mathbf{y}_{rec} = \mathbf{A}_{\mathbf{S}}(\mathbf{A}_{\mathbf{S}}^T\mathbf{A}_{\mathbf{S}})^{-1}\mathbf{A}_{\mathbf{S}}^T\mathbf{y} \tag{37}$$

where $\mathbf{S}$ is the support of indices of the selected potential endmembers at current iteration. Then the residual can be obtained as

$$\mathbf{r} = \mathbf{y} - \mathbf{y}_{rec} \tag{38}$$

$$= (\mathbf{I} - \mathbf{A}_{\mathbf{S}}(\mathbf{A}_{\mathbf{S}}^T\mathbf{A}_{\mathbf{S}})^{-1}\mathbf{A}_{\mathbf{S}}^T)\mathbf{y} \tag{39}$$

Thus $\mathbf{r}$ is orthogonal to $\mathbf{y}_{rec}$. So, the following equality holds

$$\|\mathbf{y}\|_2^2 = \|\mathbf{r} + \mathbf{y}_{rec}\|_2^2 \tag{40}$$

$$= \|\mathbf{r}\|_2^2 + \|\mathbf{y}_{rec}\|_2^2 \tag{41}$$

As $\mathbf{y}$ has unit length, the lemma can be easily obtained. □

Proof of Theorem 4.1: The proof is inspired by [44]. According to Algorithm 3, a sufficient condition for SMP to pick the spectral signatures only from $\mathbf{S}_{opt}$ at each iteration is that

$$\begin{cases} \|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} > \|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty} \\ \|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty} \leq t \end{cases} \tag{42}$$

From Lemma A.4 we can get that $\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} \leq 1$. Then a sufficient condition to obtain Eq. (42) is that

$$\|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty} < t\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} \tag{43}$$

We now produce an upper bound for $\frac{\|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty}}{\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty}}$.

Since $(\mathbf{A}_{opt}^\dagger)^T\mathbf{A}_{opt}^T = \mathbf{A}_{opt}(\mathbf{A}_{opt}^T\mathbf{A}_{opt})^{-1}\mathbf{A}_{opt}^T$ is a projection matrix to the subspace spanned by the columns of $\mathbf{A}_{opt}$ and $\mathbf{R}$ is in the subspace spanned by the columns of $\mathbf{A}_{opt}$, we have

$$\mathbf{R} = (\mathbf{A}_{opt}^\dagger)^T\mathbf{A}_{opt}^T\mathbf{R} \tag{44}$$

Using the equality above, we have

$$\frac{\|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty}}{\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty}} = \frac{\|\mathbf{A}_{err}^T(\mathbf{A}_{opt}^\dagger)^T\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty}}{\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty}} \tag{45}$$

$$\leq \|\mathbf{A}_{err}^T(\mathbf{A}_{opt}^\dagger)^T\|_{\infty,\infty} \tag{46}$$

$$= \|\mathbf{A}_{opt}^\dagger\mathbf{A}_{err}\|_{1,1} \tag{47}$$

$$= \max_{j \notin \mathbf{S}_{opt}} \|\mathbf{A}_{opt}^\dagger\mathbf{A}_j\|_1 \tag{48}$$

In the above deduction, we use Lemma A.2 and Lemma A.1, respectively. Thus, if Eq. (22) is satisfied, SMP will recover all the endmembers listed in the support $\mathbf{S}_{opt}$. □

APPENDIX B
PROOF OF THEOREM 4.2

First of all, we should introduce some other definitions that will be used during the proof to make it clearer.

Definition B.1: Given a hyperspectral data matrix $\mathbf{Y} \in R^{L\times K}$, a spectral library matrix $\mathbf{A} \in R^{L\times m}$,

$\mathbf{Y}_{opt}$ is an optimal P-term approximation of $\mathbf{Y}$ and it lies in the column range of $\mathbf{A}_{opt}$, $\mathbf{Y}_{opt} \in R^{L\times K}$.

$\mathbf{X}_{opt}$ is the optimal abundance matrix corresponding to $\mathbf{Y}_{opt}$, $\mathbf{X}_{opt} \in R^{m\times K}$, $\mathbf{Y}_{opt} = \mathbf{A}\mathbf{X}_{opt}$, $\|\mathbf{X}_{opt}\|_{\text{row-0}} = P$;

$\mathbf{S}_k$ is a subset of $\mathbf{S}_{opt}$ and it lists the first k endmembers SMP selects from $\mathbf{S}_{opt}$ after some iterations;

$\widehat{\mathbf{A}}_{opt}$ is a submatrix of $\mathbf{A}_{opt}$ and it contains the k endmembers that are indexed by $\mathbf{S}_k$, $\widehat{\mathbf{A}}_{opt} \in R^{L\times k}$;

$\widetilde{\mathbf{A}}_{opt}$ is a submatrix of $\mathbf{A}_{opt}$ and it contains the $(P-k)$ endmembers that are indexed by $\mathbf{S}_{opt} \setminus \mathbf{S}_k$, $\widetilde{\mathbf{A}}_{opt} \in R^{L\times(P-k)}$;

$\widehat{\mathbf{X}}_{opt}$ is a submatrix of $\mathbf{X}_{opt}$ and it contains the rows listed in $\mathbf{S}_k$, $\widehat{\mathbf{X}}_{opt} \in R^{k\times K}$;

$\widetilde{\mathbf{X}}_{opt}$ is a submatrix of $\mathbf{X}_{opt}$ and it contains the rows listed in $\mathbf{S}_{opt} \setminus \mathbf{S}_k$, $\widetilde{\mathbf{X}}_{opt} \in R^{(P-k)\times K}$.

Lemma B.1: Suppose after some iterations, SMP has picked k optimal endmembers listed in $\mathbf{S}_{opt}$ and $\mu_1(P) < \frac{t}{1+t}$, $M(k) < \frac{t}{1+t}$, then the following bounds are in force:

$$\|\mathbf{A}_{err}^T(\mathbf{Y}_{opt} - \mathbf{Y}_{rec})\|_{1,\infty} \leq M(k)\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty}, \tag{49}$$

$$\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} \geq (1 - M(k))\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{50}$$

where $M(k) = M(k;P) = \frac{\mu_1(P-k)}{1-\mu_1(k)}$.

Proof: The proof is inspired by [43], however there are some notable differences. Let $\mathbf{P}_k$ be the orthogonal projector onto the range of $\widehat{\mathbf{A}}_{opt}$

$$\mathbf{P}_k = \widehat{\mathbf{A}}_{opt}(\widehat{\mathbf{A}}_{opt}^T\widehat{\mathbf{A}}_{opt})^{-1}\widehat{\mathbf{A}}_{opt}^T \tag{51}$$

Since $(\mathbf{Y} - \mathbf{Y}_{opt})$ is orthogonal to the range of $\mathbf{A}_{opt}$, we have

$$\mathbf{A}_{opt}^T\mathbf{R} = \mathbf{A}_{opt}^T(\mathbf{Y} - \mathbf{Y}_{opt}) + \mathbf{A}_{opt}^T(\mathbf{Y}_{opt} - \mathbf{Y}_{rec}) \tag{52}$$

$$= \mathbf{A}_{opt}^T(\mathbf{Y}_{opt} - \mathbf{Y}_{rec}) \tag{53}$$

Then, rewrite the matrix $(\mathbf{Y}_{opt} - \mathbf{Y}_{rec})$ in the following manner:

$$\mathbf{Y}_{opt} - \mathbf{Y}_{rec} = \mathbf{Y}_{opt} - \mathbf{P}_k\mathbf{Y}_{opt} \tag{54}$$

$$= (\mathbf{I} - \mathbf{P}_k)\mathbf{Y}_{opt} \tag{55}$$

$$= (\mathbf{I} - \mathbf{P}_k)\mathbf{A}_{opt}\mathbf{X}_{opt} \tag{56}$$

$$= (\mathbf{I} - \mathbf{P}_k)\widetilde{\mathbf{A}}_{opt}\widetilde{\mathbf{X}}_{opt} \tag{57}$$

The last equality holds because $(\mathbf{I}-\mathbf{P}_k)$ is orthogonal to the range of $\widehat{\mathbf{A}}_{opt}$. We first build an upper bound for $\|\mathbf{A}_{err}^T(\mathbf{Y}_{opt}-\mathbf{Y}_{rec})\|_{1,\infty}$. Apply the triangle inequality and Lemma A.2:

$$\|\mathbf{A}_{err}^T(\mathbf{Y}_{opt}-\mathbf{Y}_{rec})\|_{1,\infty} \tag{58}$$

$$= \|\mathbf{A}_{err}^T(\mathbf{I}-\mathbf{P}_k)\widetilde{\mathbf{A}}_{opt}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{59}$$

$$\leq \|\mathbf{A}_{err}^T\widetilde{\mathbf{A}}_{opt}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} + \|\mathbf{A}_{err}^T\mathbf{P}_k\widetilde{\mathbf{A}}_{opt}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{60}$$

$$\leq (\|\mathbf{A}_{err}^T\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} + \|\mathbf{A}_{err}^T\mathbf{P}_k\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty})\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{61}$$

Then, we find an upper bound for the bracketed term above. This process takes five steps:

(1) Since each row of the matrix $(\mathbf{A}_{err}^T\widetilde{\mathbf{A}}_{opt})$ lists the inner products between a spectral signature and $(P-k)$ distinct signatures, we have $\|\mathbf{A}_{err}^T\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} \leq \mu_1(P-k)$.

(2) Use the fact that $\|\cdot\|_{\infty,\infty}$ is submultiplicative and the definition of $\mathbf{P}_k$ to see that

$$\|\mathbf{A}_{err}^T\mathbf{P}_k\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} \leq$$
$$\|\mathbf{A}_{err}^T\widehat{\mathbf{A}}_{opt}\|_{\infty,\infty}\|(\widehat{\mathbf{A}}_{opt}^T\widehat{\mathbf{A}}_{opt})^{-1}\|_{\infty,\infty}\|\widehat{\mathbf{A}}_{opt}^T\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} \tag{62}$$

(3) Using the same deduction as in Step (1), we can get that $\|\mathbf{A}_{err}^T\widehat{\mathbf{A}}_{opt}\|_{\infty,\infty} \leq \mu_1(k)$ and $\|\widehat{\mathbf{A}}_{opt}^T\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} \leq \mu_1(P-k)$.

(4) Since each column of $\mathbf{A}$ has l_2-norm unit length, the Gram matrix $(\widehat{\mathbf{A}}_{opt}^T\widehat{\mathbf{A}}_{opt})$ has a unit diagonal. Hence, we can write

$$\widehat{\mathbf{A}}_{opt}^T\widehat{\mathbf{A}}_{opt} = \mathbf{I}-\mathbf{C} \tag{63}$$

where $\|\mathbf{C}\|_{\infty,\infty} \leq \mu_1(k)$. Then we expand the inverse in a Neumann series to obtain an upper bound for $\|(\widehat{\mathbf{A}}_{opt}^T\widehat{\mathbf{A}}_{opt})^{-1}\|_{\infty,\infty}$

$$\|(\widehat{\mathbf{A}}_{opt}^T\widehat{\mathbf{A}}_{opt})^{-1}\|_{\infty,\infty} = \|(\mathbf{I}-\mathbf{C})^{-1}\|_{\infty,\infty} \tag{64}$$

$$\leq \frac{1}{1-\|\mathbf{C}\|_{\infty,\infty}} \tag{65}$$

$$\leq \frac{1}{1-\mu_1(k)} \tag{66}$$

(5) Introduce the bounds from Steps (1)-(4) into Eq. (61). We can get that

$$\|\mathbf{A}_{err}^T(\mathbf{Y}_{opt}-\mathbf{Y}_{rec})\|_{1,\infty} \tag{67}$$

$$\leq [\mu_1(P-k) + \frac{\mu_1(k)\mu_1(P-k)}{1-\mu_1(k)}]\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{68}$$

$$= \frac{\mu_1(P-k)}{1-\mu_1(k)}\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{69}$$

$$= M(k)\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{70}$$

Now we turn to produce a lower bound for $\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty}$. Invoke Eq. (53) and Eq. (57) to get

$$\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} = \|\widetilde{\mathbf{A}}_{opt}^T(\mathbf{I}-\mathbf{P}_k)\widetilde{\mathbf{A}}_{opt}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{71}$$

For simplicity, let $\mathbf{D} = \widetilde{\mathbf{A}}_{opt}^T(\mathbf{I} - \mathbf{P}_k)\widetilde{\mathbf{A}}_{opt}$. Using Lemma A.3, we can get

$$\|\mathbf{D}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \geq \|\mathbf{D}^{-1}\|_{\infty,\infty}^{-1}\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{72}$$

Write $\mathbf{D} = \mathbf{I} - (\mathbf{I} - \mathbf{D})$, and expand its inverse in Neumann series to get

$$\|\mathbf{D}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \geq (1 - \|\mathbf{I} - \mathbf{D}\|_{\infty,\infty})\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{73}$$

Recall the definition of $\mathbf{D}$ and use the triangle inequality to get

$$\|\mathbf{I} - \mathbf{D}\|_{\infty,\infty} \leq \|\mathbf{I} - \widetilde{\mathbf{A}}_{opt}^T\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} + \|\widetilde{\mathbf{A}}_{opt}^T\mathbf{P}_k\widetilde{\mathbf{A}}_{opt}\|_{\infty,\infty} \tag{74}$$

$$\leq \mu_1(P-k) + \frac{\mu_1(P-k)\mu_1(k)}{1-\mu_1(k)} \tag{75}$$

$$= \frac{\mu_1(P-k)}{1-\mu_1(k)} \tag{76}$$

The second inequality can be derived analogous with Steps (1)-(5). Introduce Eq. (76) and the definition of $\mathbf{D}$ into Eq. (73) to conclude

$$\|\widetilde{\mathbf{A}}_{opt}^T(\mathbf{I} - \mathbf{P}_k)\widetilde{\mathbf{A}}_{opt}\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \geq (1 - \frac{\mu_1(P-k)}{1-\mu_1(k)})\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{77}$$

$$= (1 - M(k))\|\widetilde{\mathbf{X}}_{opt}\|_{1,\infty} \tag{78}$$

Thus, the lemma has been proved. □

Proof of Theorem 4.2: A sufficient condition for SMP to select only optimal endmembers from $\mathbf{S}_{opt}$ is that

$$\|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty} < t\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} \tag{79}$$

We can build an upper bound for $\|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty}$ using Lemma B.1.

$$\|\mathbf{A}_{err}^T\mathbf{R}\|_{1,\infty} \tag{80}$$

$$\leq \|\mathbf{A}_{err}^T(\mathbf{Y} - \mathbf{Y}_{opt})\|_{1,\infty} + \|\mathbf{A}_{err}^T(\mathbf{Y}_{opt} - \mathbf{Y}_{rec})\|_{1,\infty} \tag{81}$$

$$\leq \|\mathbf{A}_{err}^T(\mathbf{Y} - \mathbf{Y}_{opt})\|_{1,\infty} + \frac{M(k)}{1-M(k)}\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} \tag{82}$$

Thus, if the following inequality holds, Eq. (79) can be satisfied

$$\|\mathbf{A}_{err}^T(\mathbf{Y} - \mathbf{Y}_{opt})\|_{1,\infty} + \frac{M(k)}{1-M(k)}\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} < \tag{83}$$

$$t\|\mathbf{A}_{opt}^T\mathbf{R}\|_{1,\infty} \tag{84}$$

Rearrange this relation to obtain Eq. (23). □

ACKNOWLEDGMENT

The authors would like to thank M.-D. Iordache, J. Bioucas-Dias and A. Plaza for sharing their codes for the algorithms of SUnSAL and SUnSAL-TV. The authors also thank Konstantinos E. Themelis and Mark A. Davenport for providing the codes for the BI-ICE algorithm and CoSaMP algorithm, respectively. Last but not least, the

authors sincerely thank the Associate Editor and the three anonymous reviewers for their very useful comments and suggestions which greatly improve the quality of this manuscript. Specially, the authors would like to express their gratitude to one of the reviews for providing the idea to take the proposed algorithm as a dictionary pruning algorithm to boost other sparse unmixing algorithms.

REFERENCES

[1] Peg Shippert, "Why use hyperspectral imagery?," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 4, pp. 377-380, Apirl 2004.

[2] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 17-28, 2002.

[3] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 44-57, 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, pp. 624-626, Jan. 1999.

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

[6] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354-379, 2012.

[7] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, vol. 1, pp. 11-14, 1993.

[8] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," *in Proc.* 13^{th} *Int. Conf. Appl. Geologic Remote Sens. Vancouver, BC, Canada, Apr.*, vol. 2, pp. 337-344, 1999.

[9] J. M. Nascimento and J.M. Bioucas-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.

[10] C.-I. Chang, C.-C. Wu, W. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 2804-2819, 2006.

[11] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, pp. 99-109, 1994.

[12] 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, 2004.

[13] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765-777, 2007.

[14] T. Chan, C. Chi, Y. Huang, and W. Ma, "Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4418-4432, 2009.

[15] M. Arngren, M. N. Schmidt, and J. larsen, "Bayesian nonnegative matrix factorization with volume prior for unmixing of hyperspectral images," *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*, pp. 1-6, Sept. 2009.

[16] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp. 4033-4036, July 2007.

[17] 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.

[18] V. Paul Pauca, J. Piper, and Robert J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29-47, Jul. 2006.

[19] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Hoboken, NJ: Wiley, 2009.

[20] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Adv. Neural Inform. Process. Syst.*, vol. 13, pp. 556-562, 2002.

[21] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "L1/2 Sparsity Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 19-60, Nov. 2011.

[22] Sen Jia and Yuntao Qian, "Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161-173, Jan. 2009.

[23] Xiaoqiang Lu, Hao Wu, Yuan Yuan, Pingkun Yan, and Xuelong Li, "Manifold Regularized Sparse NMF for Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sens.*, Article in Press, doi: 10.1109/TGRS.2012.2213825.

[24] 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.

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

[26] Konstantinos E. Themelis, Athanasios A. Rontogiannis, and Konstantinos D. Koutroumbas, "A Novel Hierarchical Bayesian Approach for Sparse Semisupervised Hyperspectral Unmixing," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 585-599, 2012.

[27] Y Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973-3985, Oct. 2011.

[28] X. Zhao, F. Wang, T. Huang, M. Ng, and R. Plemmons, "Deblurring and Sparse Unmixing For Hyperspectral Images," *IEEE Trans. Geosci. Remote Sens.*, Article in Press, doi: 10.1109/TGRS.2012.2227764.

[29] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484-4502, 2012.

[30] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative Sparse Regression for Hyperspectral Unmixing,"

IEEE Transactions on Geoscience and Remote Sensing, Article in Press, doi: 10.1109/TGRS.2013.2240001.

[31] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Dictionary pruning in sparse unmixing of hyperspectral Data," in *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'12), Shanghai, China*, pp. 1-4, 2012.

[32] Yonina C. Eldar and Holger Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 505-519, 2010.

[33] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.

[34] Shihao Ji, Ya Xue, and Lawrence Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, pp. 2346-2356, 2008.

[35] Michael Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[36] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein, *Introduction to Algorithms (3rd ed.)*. MIT Press and McGraw-Hill, 2009.

[37] J. A. Tropp, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[38] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to snderdetermined systems of equations," *IEEE Trans. Inform. Theory*, vol. 54, no. 11, pp. 4813-4820, Nov. 2008.

[39] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129-159, 2001.

[40] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267-288, 1996.

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

[42] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: application to hyperspectral unmixing," *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pp. 1-4, June 2010.

[43] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation: part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572-588, May 2005.

[44] J. Chen and Xiaoming Huo, "Theoretical Results on Sparse Representations of Multiple-Measurement Vectors," *IEEE Transactions on signal processing*, vol. 54, no. 12, Dec 2006.

[45] Yi Chen, Nasser M. Nasrabadi, and Trac D. Tran, "Joint sparsity for target detection," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVII*, vol. 8048, 2011.

[46] Wei Dai and O. Milenkovic, "Subspace Pursuit for Compressive Sensing Signal Reconstruction," *Information*

Theory, IEEE Transactions on, vol. 55, no. 5, pp. 2230-2249, May 2009.

[47] D. Needell , J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301-321, 2009.

[48] Mark A. Davenport, Deanna Needell, and Michael B. Wakin, "Signal Space CoSaMP for Sparse Recovery with Redundant Dictionaries," *arXiv preprint arXiv:1208.0353*, 2012.

[49] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," *J. Constr. Approx.*, vol. 13, pp. 57-98, 1997.

[50] D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197-2202, Mar. 2003.

[51] M.-D. Iordache, "A sparse regression approach to hyperspectral unmixing," *Technical University of Lisbon, PhD Thesis*, Nov. 2011.

[52] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 650-663, Mar. 2004.

[53] R. N. Clark, G. A. Swayze, A. J. Gallagher, T. V. V. King, and W. M.Calvin, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," *U.S. Geological Survey, Denver, CO, Open File Rep.*, pp. 93-592, 1993.

[54] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608-619, Mar. 2004.

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