Multi-objective based Spectral Unmixing for Hyperspectral Images ☆

Xia Xu[a], Zhenwei Shi[a,b,c,*]

[a] *Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, PR China*

[b] *State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, PR China*

[c] *Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, PR China*

Abstract

Sparse hyperspectral unmixing assumes that each observed pixel can be expressed by a linear combination of several pure spectra in a priori library. Sparse unmixing is challenging, since it is usually transformed to a NP-hard l_0 norm based optimization problem. Existing methods usually utilize a relaxation to the original l_0 norm. However, the relaxation may bring in sensitive weighted parameters and additional calculation error. In this paper, we propose a novel multi-objective based algorithm to solve the sparse unmixing problem without any relaxation. We transform sparse unmixing to a multi-objective optimization problem, which contains two correlative objectives: minimizing the reconstruction error and controlling the endmember sparsity. To improve the efficiency of multi-objective optimization, a population-based randomly flipping strategy is designed. Moreover, we theoretically prove that the proposed method is able to recover a guaranteed approximate solution from the spectral library within limited iterations. The proposed method can

☆The work was supported by the National Natural Science Foundation of China under the Grants 61273245 and 61671037, the Beijing Natural Science Foundation under the Grant 4152031, the funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under the Grant BUAA-VR-16ZZ-03, and the Fundamental Research Funds for the Central Universities under the Grant YWF-14-YHXY-028 and YWF-15-YHXY-003.

*Corresponding author: Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, PR China. Tel: +86 010 82339520; Fax: +86 010 82338798.

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

directly deal with l_0 norm via binary coding for the spectral signatures in the library. Experiments on both synthetic and real hyperspectral datasets demonstrate the effectiveness of the proposed method.

Keywords: Hyperspectral image, Sparse unmixing, Multi-objective optimization, l_0 problem, Binary coding

1. Introduction

Hyperspectral imaging technology has achieved prominent development in recent years [Bioucas-Dias et al. 2012]. Hyperspectral images usually cover hundreds of bands which can be detected simultaneously through imaging spectrometer. Based on such high spectral resolution, a lot of valuable information is extracted [Green et al. 1998, Keshava et al. 2002, Shippert. 2004, Landgrebe. 2002]. However, the spatial resolution of most hyperspectral images is low, thus observed pixels are usually mixed which means several pure materials could be found in a single pixel [Bioucas-Dias et al. 2013]. The existence of mixed pixels imposes restrictions on practical applications of hyperspectral images. Therefore, decomposition of the measured spectral signals, commonly known as hyperspectral unmixing, is important for accurate interpretation of hyperspectral image.

Hyperspectral unmixing refers to the process of separating mixed pixels into a set of constitutive spectra (endmembers) and their corresponding fractions (abundances) [Hu et al. 1999, Petrou et al. 1999, Dobigeon et al. 2008, Zhong et al. 2016]. The endmembers are assumed as an representation of the pure substances presented in the image, and the abundance fractions are their respective percentages in each pixel. The decomposition operation consists of two steps: extracting endmembers and estimating the corresponding abundances [Parra et al. 1999]. Linear mixing model (LMM) is popular to model spectral unmixing problem [Adams et al. 1986, Ma et al. 2014]. It assumes that each pixel in the hyperspectral image is linearly weighted by the endmembers existing in the pixel. Based on LMM, there has been a tremendous effort in the past decade to solve the spectral unmixing problem. Early research are mainly based on geometry [Boardman. 1993, Winter. 2003, Nascimento et al. 2005], statistics [Nascimento et al. 2005] and nonnegative matrix factorization (NMF) [Qian et al. 2011, Lu et al. 2013, Zhu et al. 2014]. The geometry-based methods assume that all the pixels lie in a high-dimensional data simplex, and endmembers are considered as the vertices of the simplex

[Boardman. 1993]. However, these methods usually assume the presence of at least one pure pixel per endmember in the image, which is a strict assumption and difficult to be satisfied [Iordache et al. 2011]. Statistics based methods transform the problem to a statistical inference form, whereas these methods are usually complex and time-consuming [Berman et al. 2004]. NMF could decompose the hyperspectral image into a product of two nonnegative matrixes, whose mathematics form is similar to that of linear spectral unmixing. However, the physical meaning of the extracted endmembers seems ambiguous [Pauca et al. 2006].

Recently, the spectral unmixing problem has been approached in a spectral library based semi-supervised fashion, known as sparse unmixing, aiming at finding an optimal subset of the library that can best model each observed mixed pixel in the scene. Spectral library is a collection of many pure spectral signatures which are known in advance. This approach often leads to a sparse solution due to the fact that the number of spectral signatures in the library is usually much larger than that of endmembers presented in the hyperspectral image. However, the objective function of sparse unmixing is a l_0 norm constrained optimization problem, which is NP-hard [Elad. 2010, Gong et al. 2016].

To solve the sparse unmixing problem, researchers have introduced many strategies, mainly including convex relaxation approaches [Eldar et al. 2010, Tropp. 2006] and greedy algorithms (GAs) [Tropp. 2004, Tropp et al. 2006]. Convex relaxation usually approximates l_0 norm to a closest convex l_1 norm, so that global optima can be achieved. Some optimization algorithm such as alternating direction method of multipliers (ADMM) [Esser. 2009, Yang et al. 2010], are used to solve the problem. In the paper [Iordache et al. 2014], sparse unmixing problem was extended to a joint sparse regression form which aims at finding a few nonzero lines in library. In the paper [Iordache et al. 2012, Feng et al. 2014, Feng et al. 2016, Feng et al. 2016], spatial information of hyperspectral images was taken into account to improve the unmixing accuracy. In view of the fact that some materials in spectral library are known to exist in the scene, the spectral a priori information was incorporated into the sparse unmixing process in [Tang et al. 2015]. However, convex relaxation based methods can only provide a limited approximation to the l_0 norm regularized sparse unmixing problem. Greedy algorithms try to find the optimal solution heuristically through making best decision in each iteration. In the paper [Iordache et al. 2011], typical orthogonal matching pursuit (OMP) was used to solve the sparse unmixing problem, where the

spectral signatures most correlated with current residual were selected to update the solution. However, the high correlation of spectral library limits its performance. To overcome the deficiency of OMP, subspace matching pursuit (SMP) [Shi et al. 2014] was proposed based on simultaneous orthogonal matching pursuit. In [Tang et al. 2014], the regularized simultaneous forward-backward greedy algorithm (RSFoBa) was proposed, which integrated a backward greedy step to the forward greedy step. In this case, incorrect extraction made in earlier forward steps can be removed and candidate endmembers can be reduced. However, the number of endmembers extracted by greedy algorithms is usually unstable. Bayesian approach was also proposed to solve sparse unmixing. In [Themelis et al. 2012], a sparsity assumption and nonnegativity constraint based hierarchical Bayesian approach was presented, which can provide the sparse solution without necessarily tuning any parameters. But it might be much more complex than the convex relaxation approaches and GAs.

In this paper, we propose a novel multi-objective based sparse unmixing (MOSU) for hyperspectral data, by transforming it to a bi-objective optimization problem. Multi-objective optimization refers to solving problems of more than one objective, which often conflicts against each other [Deb et al. 2001, Sindhya et al. 2013]. This property is precisely in line with the two objectives in sparse unmixing. The motivation of our study is the excellent performance of the subset selection by Pareto optimization (POSS) [Qian et al. 2015]. POSS treats subset selection as a multi-objective optimization problem and uses Pareto optimization to update the candidate solutions. Based on POSS, we express sparse hyperpsectral unmixing as a bi-objective optimization. The reconstruction error and sparsity of the solution are treated as two objectives that can be optimized simultaneously under our framework. Then, the l_0 norm based sparse unmixing can be solved directly without any relaxation. However, POSS is time-consuming, since the convergence speed is quite slow. In MOSU, we improve the computational efficiency of POSS by extending it to a population-based method. Since POSS cannot directly handle the population problem, we integrate the improved method into the framework of non-dominated sorting genetic algorithm-II (NSGA-II) [Deb et al. 2002]. Original POSS adopted a randomly flipping strategy to select subset from a large set of variables, and only one offspring is generated in each iteration. Here, we replace the single updating strategy by a population-based strategy. By generating more offsprings in each iteration, our algorithms can find solutions with lower computation work,

compared with POSS. A theoretical proof about the maximum iterations of our method to obtain a guaranteed approximate solution, is further presented. The main contributions of this paper are listed below:

- We introduce a novel multi-objective optimization based method for hyperspectral sparse unmixing, by transforming sparse unmixing to a bi-objective optimization problem. Then the l_0 norm constrained unmixing problem is able to be handled directly without relaxation.
- To improve the computational efficiency of the POSS, we extend the POSS to a population-based algorithm, where several offsprings are generated in each main iteration.
- Furthermore, we have theoretically present the maximum iterations of MOSU under the same condition of POSS.

The rest of this paper is organized as follows. Section 2 introduces some related works concerned with the LMM, POSS and NSGA-II. In Section 3, we present the MOSU and give some theoretical analysis. Experimental results are shown in Section 4. Finally, we conclude this paper in Section 5.

2. Related works

In this section, some related works are presented. We first give a brief introduction to the linear sparse unmixing model. Then, POSS and NSGA-II, which lay a foundation for our proposed algorithm, are reviewed.

2.1. Linear sparse unmixing model

In hyperspectral images, a pixel usually contains several different materials. However, only a single spectrum is recorded to express the pixel, resulting in the phenomenon of mixed observed spectra. In linear sparse unmixing model, each pixel in a hyperspectral image can be approximated as a linear combination of several pure spectral signatures in a spectral library:

$$\mathbf{y} = \sum_{i=1}^{m} \mathbf{a}_i x_i + \mathbf{n} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{r\times 1}$ is the measured spectrum of a mixed pixel with r bands, $\mathbf{A} \in \mathbb{R}^{r\times m}$ is the spectral library with m spectral signatures, $\mathbf{x} \in \mathbb{R}^{m\times 1}$

is the abundance vector with regard to the library $\mathbf{A}$, and $\mathbf{n} \in \mathbb{R}^{r\times 1}$ is the error term. Considering the physical meaning of abundance fractions, two constrains should be imposed to the sparse unmixing model: the abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC):

$$\text{ANC} : x_i \geq 0 \tag{2}$$

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

However, ASC is prone to strong criticism for real hyperspectral images. It should be extended to a so-called generalized sum-to-one constraint (generalized ASC) which could be automatically imposed by the nonnegativity of sources [Iordache et al. 2011]. As a result, the optimization problem of sparse unmixing can be written as:

$$\begin{aligned} &\min_{\mathbf{x}} \|\mathbf{x}\|_0 \\ &s.t. \ \ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \delta, \mathbf{x} \geq 0 \end{aligned} \tag{4}$$

where $\|\mathbf{x}\|_0$ denotes the number of non-zero elements in $\mathbf{x}$, $\delta > 0$ is the tolerance of reconstruction error. Suppose the hyperspectral image contains n pixels, then Eq. (1) can be written in a matrix form:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{5}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n] \in \mathbb{R}^{r\times n}$ denotes the whole hyperspectral image data, $\mathbf{y}_i \in \mathbb{R}^{r\times 1}$ is the ith pixel data, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ denotes the abundance matrix, $\mathbf{x}_i \in \mathbb{R}^{m\times 1}$ is the abundance vector of the ith pixel, and $\mathbf{N}$ is the noise matrix. Based on this model, the sparse unmixing problem can be expressed by

$$\begin{aligned} &\min_{\mathbf{X}} \|\mathbf{X}\|_{row-0} \\ &s.t. \ \ \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F \leq \delta, \mathbf{X} \geq 0 \end{aligned} \tag{6}$$

where $\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F = \sqrt{trace[(\mathbf{Y} - \mathbf{A}\mathbf{X})(\mathbf{Y} - \mathbf{A}\mathbf{X})^T]}$ is the Frobenius norm of $\mathbf{X}$, $\|\mathbf{X}\|_{row-0}$ denotes the number of nonzero rows (sparsity) in abundance matrix $\mathbf{X}$. Unfortunately, due to the combination and non-smooth properties of l_0 norm, the optimization problem is NP-hard to solve.

2.2. The POSS method

The subset selection by Pareto optimization (POSS) is an effective multi-objective based subset selection method, which aims at solving problems as feature selection, sparse regression and dictionary learning [Qian et al. 2015]. Here, we first give a brief introduction to the multi-objective optimization, and then explain how to apply POSS in sparse regression.

2.2.1. Multi-objective optimization

Multi-objective optimization is a class of algorithms for searching optimal solution of problems with several objectives [Li et al. 2013, Xu et al. 2014]. Mathematically, multi-objective problem can be described as follows [Eckart et al. 1999]:

$$\begin{aligned} &\min f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_N(\mathbf{x})] \\ &s.t. \quad g_i(\mathbf{x}) \leq 0, i = 1, 2, ..., p \\ &\qquad h_j(\mathbf{x}) = 0, j = 1, 2, ..., q \end{aligned} \tag{7}$$

where $\mathbf{x}$ is the decision variable, $f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_N(\mathbf{x})]$ denotes the set of objective functions, $g_i(\mathbf{x})$ and $h_j(\mathbf{x})$ are the constraint functions. Different from traditional single-objective optimization, the objectives in multi-objective optimization usually conflict with each other. Thus, there may not exist a solution optimal to all the objectives. Generally, a compromise should be adopted to make all the objectives as optimal as possible. A common strategy for multi-objective optimization is giving an optimal solution set, known as Pareto optimal. As discussed in [Eckart et al. 1999], considering two decision variables $\mathbf{x}_1$ and $\mathbf{x}_2$ for a minimization problem, $\mathbf{x}_1$ dominates $\mathbf{x}_2$ (i.e. $\mathbf{x}_1 \succ \mathbf{x}_2$) if

$$\forall i \in \{1, 2, ..., N\} : f_i(\mathbf{x}_1) \leq f_i(\mathbf{x}_2) \wedge \exists j \in \{1, 2, ..., N\} : f_j(\mathbf{x}_1) < f_j(\mathbf{x}_2) \tag{8}$$

2.2.2. POSS in sparse regression

Considering the sparse regression problem:

$$\begin{aligned} &\min_{\mathbf{s}} f(\mathbf{s}) \\ &s.t. \quad |\mathbf{s}| \leqslant k \end{aligned} \tag{9}$$

where $f(\mathbf{s})$ is the criterion function, $\mathbf{s}$ is a binary vector of all the columns of $\mathbf{V}$ (each column is given a label 0 or 1), $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m]$ can be seen

as a library in sparse regression problems (if an atom $\mathbf{v}_i \in \mathbf{V}$ is selected, the corresponding position in $\mathbf{s}$ is set as 1), $|\mathbf{s}|$ is the cardinality of the set $\mathbf{s}$, $\mathbf{V_s}$ is a subset of $\mathbf{V}$ with $|\mathbf{s}|$ atoms, and k is a positive integer to control the sparsity of the solution. In POSS, the constrained optimization problem of Eq. (9) is treated as a bi-objective optimization. The two objectives are the criterion function and the sparsity constraint. Then Eq. (9) can be described as

$$\min_{\mathbf{s}} f(\mathbf{s}) = [f_1(\mathbf{s}), f_2(\mathbf{s})] \tag{10}$$

where

$$\begin{aligned} f_1(\mathbf{s}) &= \begin{cases} +\infty, & |\mathbf{s}| = 0 \ or \ \ \geq 2k \\ f(\mathbf{s}), & otherwise \end{cases} \\ f_2(\mathbf{s}) &= |\mathbf{s}| \end{aligned} \tag{11}$$

POSS begins with an empty set, where each decomposed column is labeled as 0. An archive P is used to store the candidate solutions which are updated iteration by iteration based on Pareto optimization. In each main iteration, a single candidate is selected from P and updated through a randomly flipping operation where each bit (0 or 1) is flipped with a probability $1/m$ or remains unchanged with a probability $1-1/m$. Then P is updated by comparing the new generated solution with all the current candidate solutions in P based on Pareto optimization. Moreover, POSS has been proved that it can find a set $\mathbf{s}$ with the condition in Theorem 1 [Qian et al. 2015] within $2ek^2m$ main iterations.

Although POSS is an useful method to deal with sparse regression problem, it is computational complex, especially when the endmember number is large.

2.3. NSGA-II

NSGA-II is a population-based evolutionary strategy. The comparison strategy in NSGA-II is effective on population-based evolutionary algorithms. The main operations of NSGA-II include non-dominated sorting, crowding distance calculating, tournament selection and genetic algorithm [Siinivas et al. 1994]. Some details are described below:

Non-dominated sorting: For an individual $\mathbf{s}$, the number of individuals that dominate it is $n_{\mathbf{s}}$ (individuals stored in $N_{\mathbf{s}}$), and the number of individuals it dominates is $s_{\mathbf{s}}$ (individuals stored in $S_{\mathbf{s}}$). Individuals with $n_{\mathbf{s}} = 0$

are added to the first front. For each individual $\mathbf{s}_1$ in the first front and each individual $\mathbf{s}'$ in $S_{\mathbf{s}_1}$, $n_{\mathbf{s}'}$ is decremented by 1. If $n_{\mathbf{s}'} = 0$, the front of individual $\mathbf{s}'$ would be set as 2. The other fronts are determined by the same way. Finally, all individuals are sorted to several fronts.

Crowding distance calculating: Individuals in the same front are ranked by their crowding distances. In each front, crowding distance is set to an infinite value for boundary individuals with maximal and minimal fitness value respectively. For the other individuals, it is calculated as below:

$$d_{\mathbf{s}} = \sum_{i=1}^{l} \frac{f_i(\mathbf{s}+1) - f_i(\mathbf{s}-1)}{f_i^{max} - f_i^{min}} \tag{12}$$

where $d_{\mathbf{s}}$ is the crowding distance of the individual $\mathbf{s}$, $f_i(\mathbf{s}+1)$ and $f_i(\mathbf{s}-1)$ are the ith objective values for the two adjacent individuals of $\mathbf{s}$ respectively, f_i^{max} and f_i^{min} are the maximal and minimal fitness values, l is the number of objectives.

Tournament selection: In each main iteration, several individuals (parents) are selected to be updated. For the selection of each parent, two individuals are selected randomly and the individual with better performance is selected based on non-dominated sorting and their crowding distances.

Genetic algorithm: Crossover and mutation of genetic algorithm are conducted to generate offsprings. Then the candidate solutions are updated based on non-dominated sorting and crowding distance.

3. Multi-optimization based spectral unmixing for hyperspectral image

Motivated by the managing ability of POSS for sparse regression problems, we introduce it to solve the sparse unmixing problem of hyperspectral images. The randomly flipping strategy in POSS is suitable to handle the sparsity of the selected endmembers. Considering the particularity of sparse regression and the high efficiency of population-based strategy, we improve POSS to a population-based method, aiming at obtaining a faster convergence speed. In this case, the comparison strategy (Pareto optimization) in POSS would be unsuitable. Therefore, we involve our improved POSS algorithm in the NSGA-II framework, so as to take advantage of its comparison strategy: non-dominated sorting and crowding distance.

In this section, we introduce our proposed multi-optimization based sparse unmixing algorithm in detail. Then, we give a theoretical analysis of the proposed algorithm.

3.1. Multi-optimization based spectral unmixing algorithm

In MOSU, the spectral unmixing problem is transformed to a bi-objective optimization problem as the form of Eq. (10) and Eq. (11), the two objectives are the reconstruction error and the sparsity of the solution. It can be described as:

$$\min_{\mathbf{s}} f(\mathbf{s}) = [f_1(\mathbf{s}), f_2(\mathbf{s})] \tag{13}$$

where

$$\begin{aligned} f_1(\mathbf{s}) &= \begin{cases} a\ large\ value, & |\mathbf{s}| = 0\ or\ \geq 2k \\ \|\mathbf{Y} - \mathbf{A}_\mathbf{s}\mathbf{X}\|_F, & otherwise \end{cases} \\ f_2(\mathbf{s}) &= |\mathbf{s}| \end{aligned} \tag{14}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m]$ is the spectral library, k denotes the number of active endmembers in the image, $\mathbf{s}$ is a binary vector of the spectral signatures, $\mathbf{A}_\mathbf{s}$ is a sub-library of $\mathbf{A}$ with $|\mathbf{s}|$ atoms. The endmember number k should be known in advance. In MOSU, we use HySime [Bioucas-Dias et al. 2008] to give it an automatic estimation. For a solution $\mathbf{s}$, if the ith atom in $\mathbf{s}$ is 1, the corresponding $\mathbf{a}_i$ is selected as an endmember. $f_1(\mathbf{s})$ is set to a large value when $|\mathbf{s}| = 0$ or $|\mathbf{s}| \geq 2k$, which is infinite in POSS. The difference here is used to adapt the crowding distance calculating. If f_1^{min} and f_1^{max} are infinite, denominator in Eq. (12) will be imponderable and the crowding distance will be meaningless. The whole process of MOSU for hyperspectral image unmixing is shown in Algorithm 1.

MOSU contains six main steps. The first five steps are conducted to extract endmembers, and the last one aims at calculating the abundance fractions for the scene.

Step 1: Initialize the population randomly and then sort them. The initialized population can be seen as an initial candidate solution set. It contains several independent binary vectors (individuals). Then the non-dominated sorting is performed to the population. Thus all the individuals are sorted to several graduated fronts.

Step 2: Compute the crowding distance for each individual based on Eq. (12) and rank the whole population further. Individuals with maximal and

Algorithm 1 The flowchart of MOSU for hyperspectral image unmixing

Initialize:

Decompose the spectral library as $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_m]$;

The randomly initialized population (initial candidate solutions) $P_0 = \{\mathbf{s}\}$ with *pop* individuals (each bit is initialized as 1 with a probability 1/m and 0 with a probability 1-1/m);

The number of parents individuals $pop' = pop/2$ and population P';

The number of main iterations *gen*.

Process:

while t is smaller than the maximum iteration number *gen* **do**

Non-Dominated sorting:

For each $\mathbf{s} \in P_t$, determine the corresponding $N_\mathbf{s}$ and $S_\mathbf{s}$.

If $N_\mathbf{s} = \emptyset$, set $rank_\mathbf{s} = 1$ and $F_1 = F_1 \cup \mathbf{s}$ (F_1 is the first front).

while front $F_i \neq \emptyset$ **do**

For each $\mathbf{s} \in F_i$, $\mathbf{s}' \in S_\mathbf{s}$, set $n_{\mathbf{s}'} = n_{\mathbf{s}'} - 1$;

If $n_{\mathbf{s}'} = 0$, set $rank_{\mathbf{s}'} = i + 1$, $F_{i+1} = F_{i+1} \cup \mathbf{s}'$;

end while

Crowding distance calculating:

while front F_i **do**

For each individual $\mathbf{s}$, calculate the crowding distance $d_\mathbf{s} = \frac{f_1(\mathbf{s}+1) - f_1(\mathbf{s}-1)}{f_1^{max} - f_1^{min}} + \frac{f_2(\mathbf{s}+1) - f_2(\mathbf{s}-1)}{f_2^{max} - f_2^{min}}$ ($d_\mathbf{s}$ is set to a large value if $\mathbf{s}$ is a boundary individual);

Sort the individuals in F_i in a descending order of their crowding distance.

Set $i = i + 1$.

end while

Tournament selection:

while the number of parent individuals is smaller than pop' **do**

Select two individuals $\mathbf{s}_i$ and $\mathbf{s}_j$ in the whole population randomly;

If $(rank_{\mathbf{s}_i} < rank_{\mathbf{s}_j}) | (rank_{\mathbf{s}_i} = rank_{\mathbf{s}_j} \wedge d_{\mathbf{s}_i} > d_{\mathbf{s}_j})$, $\mathbf{s}_i$ is selected as a parent individual, $P' = P' \cup \mathbf{s}_i$;

Elseif $(rank_{\mathbf{s}_i} > rank_{\mathbf{s}_j}) | (rank_{\mathbf{s}_i} = rank_{\mathbf{s}_j} \wedge d_{\mathbf{s}_i} < d_{\mathbf{s}_j})$, $\mathbf{s}_j$ is added to the parent individual, $P' = P' \cup \mathbf{s}_j$.

end while

Offspring generating (population-based randomly flipping):

for each individual $\mathbf{s} \in P'$ **do**

Update $\mathbf{s}$ through flipping each bit with a probability $1/m$ and remaining it unchanged with a probability $1 - 1/m$;

Compute $\mathbf{X} = (\mathbf{A}_\mathbf{s}^T \mathbf{A}_\mathbf{s})^{-1} \mathbf{A}_\mathbf{s}^T \mathbf{Y}$;

Compute $f_1(\mathbf{s}) = \|\mathbf{Y} - \mathbf{A}_\mathbf{s}\mathbf{X}\|_F$ and $f_2(\mathbf{s}) = |\mathbf{s}|$.

end for

Candidate solutions updating:

Merge the current candidate solutions and offsprings through putting the two populations together to a single population. After the merging operation, the new population contains $1.5 * pop$ individuals.

The first *pop* individuals will be selected based on non-dominated sorting and their crowding distance, then the candidate solutions are updated to P_{t+1} (the whole population contains *pop* individuals).

end while

Select final solution from the candidate solutions based on k (we determine final solutions whose endmember number is k), and determine the final extracted endmembers ($\mathbf{A}_{\mathbf{s}^*}$).

Inversion (abundance estimation):

Compute abundance fractions for the scene using nonnegative least squares algorithm based on the original hyperspectral data and spectral library:

$\mathbf{X}^* = arg\min_\mathbf{X} \|\mathbf{Y} - \mathbf{A}_{\mathbf{s}^*}\mathbf{X}\|, s.t. \mathbf{X} \geq 0$

minimal objective function values are assigned a large value, so that they can always be selected.

Step 3: Select parent population based on a tournament selection strategy. Individuals with relatively good performance can be selected.

Step 4: Update the parent individuals to offsprings based on a population-based randomly flipping. Each bit of parent individuals is flipped with a probability $1/m$, and remains unchanged with a probability $1 - 1/m$. Thus the number of flipped bits for each individual is random. Based on this, a sufficient exploration for the spectral signatures in the prior library can be ensured. The two objective values of these offsprings are also calculated in this step. The population-based randomly flipping strategy is further illustrated in Fig. 1.

Step 5: Mix the current candidate solutions and offsprings together. All of them are compared to update the candidate solutions whose number is the same with the original population.

Up to the fifth step, candidate solution in recent iteration are determined and the population is updated. This procedure needs to be conducted many times until the number of maximal iterations is reached. Then the final solution is selected from the final population. According to the endmember number estimated by HySime, only candidate solution whose non-zero elements is k can be selected as final solution. At last, the abundance estimation step (*Step 6*) is conducted to calculate the corresponding abundance fractions. In view of the nonnegative constraint of the abundance fractions, a nonnegative least squares algorithm (*lsqnonneg* function in MATLAB) is used.

For further describing MOSU, the overall procedure is illustrated in Fig. 2.

As mentioned previously, the proposed MOSU is an improvement of POSS, nevertheless, they are quite different: Firstly, in POSS, one of the candidate solutions in archive P is selected to be updated in each iteration. However, MOSU is a population-based algorithm, where several candidate solutions can be selected to be updated. Secondly, the updated solutions (offsprings) in MOSU can compare with not only the current candidate solutions but also the other offsprings to update the candidates. While the single offspring in POSS can only compare with the current candidate solutions. Lastly, the parent individual in POSS is selected randomly. However, parents in MOSU is selected based on a tournament strategy, where individuals with relatively better performance (elites) are selected to be updated.

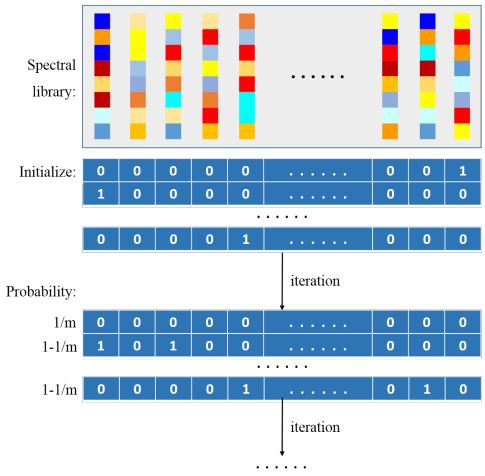

Fig. 1: The population-based randomly flipping process of MOSU. A population is initialized randomly and each bit is flipped with a probability $1/m$.

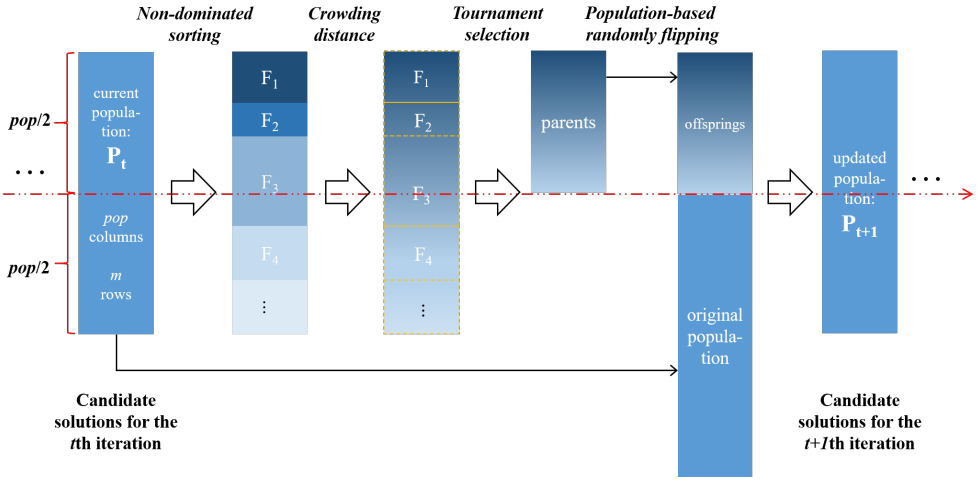

Fig. 2: The overall procedure of MOSU.

The elite selection can accelerate the algorithm and benefit to capture good solutions.

In summation, there are two main advantages of MOSU. On the one hand, the spectral unmixing problem of hyperspectral image is transformed to a bi-objective optimization problem. In this case, the l_0 norm based sparse unmixing can be handled directly without relaxation. On the other hand, a faster convergence speed can be obtained compared with POSS. MOSU improves POSS to a population-based algorithm. Hence in this case, the relatively better individuals can be explored.

3.2. Theoretical analysis

In this section, we give some theoretical analysis for MOSU.

Defination 3.1:(Submodularity Ratio [Das et al. 2011]) Let f be a non-negative set function. The submodularity ratio of f with respect to a set U and a parameter $k \geq 1$ is

$$\gamma_{U,k}(f) = \min_{L \subseteq U, S:|S| \leq k, S \cap L = \emptyset} \frac{\sum_{x \in S} f(L \cup \{x\}) - f(L)}{f(L \cup S) - f(L)} \tag{15}$$

Defination 3.2:([Das et al. 2011]) For a hyperspectral image $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n]$, finding a subset with sparsity k from a known library $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n]$ is equal to maximizing

$$F(D) = \sum_{i=1}^{n} \max_{\mathbf{A_s} \subset A, |\mathbf{s}|=k} R^2_{\mathbf{y}_i, \mathbf{A_s}} \tag{16}$$

where $R^2_{\mathbf{y}_i,\mathbf{A_s}} = (\mathrm{Var}(\mathbf{y}_i) - \mathrm{MSE}_{\mathbf{y}_i,\mathbf{A_s}})/\mathrm{Var}(\mathbf{y}_i)$ and $\mathrm{MSE}_{\mathbf{y}_i,\mathbf{A_s}} = \mathbb{E}[\|\mathbf{y}_i - \sum_{i=1}^{m} \mathbf{a}_i x_i\|^2]$ [Qian et al. 2015]. In this case, for each pixel data $\mathbf{y}$, the submodularity ratio for sparse unmixing can be defined as

$$\gamma_{U,k} = \min_{L \subseteq U, \mathbf{A_s}: |\mathbf{s}| \leq k, \mathbf{A_s} \cap L = \emptyset} \frac{\sum_{\mathbf{a} \in \mathbf{A_s}} (R^2_{\mathbf{y}, L \cup \{\mathbf{a}\}} - R^2_{\mathbf{y},L})}{R^2_{\mathbf{y},\mathbf{A_s} \cup L} - R^2_{\mathbf{y},L}} \tag{17}$$

All the following analyses are described in the pixel-wise case.

Lemma 3.1: Let $\mathbf{A}^*_{\mathbf{s}_k} \subset \mathbf{A}$ be the optimal solution, $\mathrm{OPT} = R^2_{\mathbf{y},\mathbf{A}^*_{\mathbf{s}_k}}$ be the optimal value of $\max_{\mathbf{s}:|\mathbf{s}|=k} R^2_{\mathbf{y},\mathbf{A_s}}$, $\mathbf{A}^i_{\mathbf{s}} \subset \mathbf{A}$ be the sub-library for any individual $\mathbf{s}^i$ in P_i chosen by MOSU in the first i iterations, $\overline{\mathbf{A}^i_{\mathbf{s}}} = \mathbf{A}^*_{\mathbf{s}_k} - \mathbf{A}^i_{\mathbf{s}}$ correspond to a binary code $\overline{\mathbf{s}^i}$, and $\mathbf{A}^{i\prime}_{\mathbf{s}} = \{\mathbf{a}' = \mathrm{Res}(\mathbf{a}, \mathbf{A}^i_{\mathbf{s}}) = \mathbf{a} -$

$\sum_{\mathbf{a}_i \in \mathbf{A}_\mathbf{s}^i} \mathbf{a}_i x_i \mid \mathbf{a} \in \overline{\mathbf{A}_\mathbf{s}^i}\}$ be the residual of $\mathbf{a} \in \overline{\mathbf{A}_\mathbf{s}^i}$ conditioned on $\mathbf{A}_\mathbf{s}^i$. There exists one variable $\hat{\mathbf{a}} \in \mathbf{A} - \mathbf{A}_\mathbf{s}^i$ such that

$$R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^i \cup \{\hat{\mathbf{a}}\}} - R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^i} \geq \frac{\gamma_{\emptyset,k}}{k} \cdot (\mathrm{OPT} - R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^i}) + \frac{(\mid \overline{\mathbf{s}^i} \mid -\gamma_{\emptyset,k})}{k} \cdot R^2_{\mathbf{y},\emptyset} \tag{18}$$

Proof. Using Lemmas 2.3 and 2.4 in [Das et al. 2008], we can easily drive that

$$R^2_{\mathbf{y},\overline{\mathbf{A}_\mathbf{s}^i} \cup \mathbf{A}_\mathbf{s}^i} = R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^i} + R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} \geq R^2_{\mathbf{y},\mathbf{A}^*_{\mathbf{s}_k}} = \mathrm{OPT} \tag{19}$$

Thus

$$R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} \geq \mathrm{OPT} - R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^i} \tag{20}$$

Based on *Defination 3.2* and $\mid \overline{\mathbf{s}^i} \mid = \mid \mathbf{s}^{i'} \mid \leq k$, we get

$$\begin{aligned} \gamma_{\emptyset,k} &= \min_{\mathbf{A}_\mathbf{s}^{i'}:|\mathbf{s}^{i'}| \leq k} \frac{\sum_{\mathbf{a}' \in \mathbf{A}_\mathbf{s}^{i'}} (R^2_{\mathbf{y},\emptyset \cup \{\mathbf{a}'\}} - R^2_{\mathbf{y},\emptyset})}{R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'} \cup \emptyset} - R^2_{\mathbf{y},\emptyset}} \\ &= \min_{\mathbf{A}_\mathbf{s}^{i'}:|\mathbf{s}^{i'}| \leq k} \frac{\sum_{\mathbf{a}' \in \mathbf{A}_\mathbf{s}^{i'}} (R^2_{\mathbf{y},\mathbf{a}'} - R^2_{\mathbf{y},\emptyset})}{R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} - R^2_{\mathbf{y},\emptyset}} \end{aligned} \tag{21}$$

Thus

$$\sum_{\mathbf{a}' \in \mathbf{A}_\mathbf{s}^{i'}} R^2_{\mathbf{y},\mathbf{a}'} - \mid \overline{\mathbf{s}^i} \mid \cdot R^2_{\mathbf{y},\emptyset} \geq \gamma_{\emptyset,k} \cdot (R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} - R^2_{\mathbf{y},\emptyset}) \tag{22}$$

and

$$\begin{aligned} \sum_{\mathbf{a}' \in \mathbf{A}_\mathbf{s}^{i'}} R^2_{\mathbf{y},\mathbf{a}'} &\geq \gamma_{\emptyset,k} \cdot (R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} - R^2_{\mathbf{y},\emptyset}) + \mid \overline{\mathbf{s}^i} \mid \cdot R^2_{\mathbf{y},\emptyset} \\ &= \gamma_{\emptyset,k} \cdot R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} + (\mid \overline{\mathbf{s}^i} \mid -\gamma_{\emptyset,k}) \cdot R^2_{\mathbf{y},\emptyset} \end{aligned} \tag{23}$$

Let $\hat{\mathbf{a}}' = arg\max_{\mathbf{a}' \in \mathbf{A}_\mathbf{s}^{i'}} R^2_{\mathbf{y},\mathbf{a}'}$, then we get that

$$\begin{aligned} R^2_{\mathbf{y},\hat{\mathbf{a}}'} &\geq \frac{\gamma_{\emptyset,k}}{\mid \overline{\mathbf{s}^i} \mid} \cdot R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} + \frac{(\mid \overline{\mathbf{s}^i} \mid -\gamma_{\emptyset,k})}{\mid \overline{\mathbf{s}^i} \mid} \cdot R^2_{\mathbf{y},\emptyset} \\ &\geq \frac{\gamma_{\emptyset,k}}{k} \cdot R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^{i'}} + \frac{(\mid \overline{\mathbf{s}^i} \mid -\gamma_{\emptyset,k})}{k} \cdot R^2_{\mathbf{y},\emptyset} \\ &\geq \frac{\gamma_{\emptyset,k}}{k} \cdot (\mathrm{OPT} - R^2_{\mathbf{y},\mathbf{A}_\mathbf{s}^i}) + \frac{(\mid \overline{\mathbf{s}^i} \mid -\gamma_{\emptyset,k})}{k} \cdot R^2_{\mathbf{y},\emptyset} \end{aligned} \tag{24}$$

Let $\hat{\mathbf{a}} \in \overline{\mathbf{A}_{\mathbf{s}}^i}$ corresponds to $\hat{\mathbf{a}}'$ with $\hat{\mathbf{a}}' = \mathrm{Res}(\hat{\mathbf{a}}, \mathbf{A}_{\mathbf{s}}^i)$ and use Lemma 2.4 in [Das et al. 2008], we get

$$R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}^i \cup \{\hat{\mathbf{a}}\}} - R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}^i} = R^2_{\mathbf{y},\hat{\mathbf{a}}'} \geq \frac{\gamma_{\emptyset,k}}{k} \cdot (\mathrm{OPT} - R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}^i}) + \frac{(\mid \overline{\mathbf{s}^i} \mid -\gamma_{\emptyset,k})}{k} \cdot R^2_{\mathbf{y},\emptyset} \quad (25)$$

□

Theorem 3.1: Let J_{max} be the maximum value of $j \in [0, k]$, then in the current population P_i, there exists an individual $\mathbf{s}$ in P_i with subset $\mathbf{A}_{\mathbf{s}}$ meets $\mid \mathbf{s} \mid \leq j$ and

$$\begin{aligned} R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}} \geq & [1 - (1 - \frac{\gamma_{\emptyset,k}}{k})^j] \cdot \mathrm{OPT} + (1 - \frac{\gamma_{\emptyset,k}}{k})^j \cdot R^2_{\mathbf{y},\emptyset} \\ & + \sum_{l=1}^{j} (1 - \frac{\gamma_{\emptyset,k}}{k})^{l-1} \frac{\mid \overline{\mathbf{s}_{j-l}} \mid -\gamma_{\emptyset,k}}{k} \cdot R^2_{\mathbf{y},\emptyset} \end{aligned} \quad (26)$$

where $\overline{\mathbf{s}_{j-l}}$ is the binary code corresponds to the sub-library $\overline{\mathbf{A}_{\mathbf{s}_{j-l}}} = \mathbf{A}^*_{\mathbf{s}_k} - \mathbf{A}_{\mathbf{s}_{j-l}}$, and $\mathbf{s}_{j-l}$ has $j - l$ non-zero elements.

Proof. The proof is inspired by [Qian et al. 2015]. It can be proved based on mathematical induction.

The initial value of J_{max} is 0 or 1, which is qualified for the above condition. Assume that currently $J_{max} = j < k$. Let $\mathbf{s}$ be the corresponding solution with subset $\mathbf{A}_{\mathbf{s}}$ which meet the above condition. Based on *Lemma 3.1*, there exists a new solution $\mathbf{s}_1$ satisfying

$$R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}_1}} - R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}} \geq \frac{\gamma_{\emptyset,k}}{k} \cdot (\mathrm{OPT} - R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}}) + \frac{(\mid \overline{\mathbf{s}} \mid -\gamma_{\emptyset,k})}{k} \cdot R^2_{\mathbf{y},\emptyset} \quad (27)$$

Then

$$\begin{aligned} R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}_1}} \geq & [1 - (1 - \frac{\gamma_{\emptyset,k}}{k})^{j+1}] \cdot \mathrm{OPT} + (1 - \frac{\gamma_{\emptyset,k}}{k})^{j+1} \cdot R^2_{\mathbf{y},\emptyset} \\ & + \sum_{l=1}^{j+1} (1 - \frac{\gamma_{\emptyset,k}}{k})^{l-1} \frac{\mid \overline{\mathbf{s}_{j+1-l}} \mid -\gamma_{\emptyset,k}}{k} \cdot R^2_{\mathbf{y},\emptyset} \end{aligned} \quad (28)$$

where $\mid \mathbf{s}_1 \mid = \mid \mathbf{s} \mid + 1 \leq j + 1$. After including $\mathbf{s}_1$ into the population, $J_{max} = j + 1$. Finally, the theorem is proved. □

Theorem 3.2: For MOSU, assume parents number is half of the original population, the maximal number of main iterations to find a solution $\mathbf{s}$ with $|\mathbf{s}| \leq k$ and $R^2_{\mathbf{y},\mathbf{A}_{\mathbf{s}}^k} \geq [1 - (1 - \frac{\gamma_{\emptyset,k}}{k})^k] \cdot \mathrm{OPT} + (1 - \frac{\gamma_{\emptyset,k}}{k})^k \cdot R^2_{\mathbf{y},\emptyset} + \sum_{l=1}^{k} (1 - \frac{\gamma_{\emptyset,k}}{k})^{l-1} \frac{|\overline{\mathbf{s}^i_{k-l}}| - \gamma_{\emptyset,k}}{k} \cdot R^2_{\mathbf{y},\emptyset}$ (under the condition in *Theorem 3.1*), is $1.5kem$.

Proof: J_{max} cannot decrease according to *Theorem 3.1* and the non-dominated sorting process in Algorithm 1. Thus, as parents number is half of the original population, J_{max} can increase by at least 1 in one iteration with a probability:

$$\frac{0.5pop}{pop + 0.5pop} \cdot \frac{1}{m} \cdot (1 - \frac{1}{m})^{m-1} \geq \frac{2}{3em} \tag{29}$$

where m is the size of spectral library, pop is the number of the original population and $0.5pop$ is the number of parent individuals generated in the tournament selection process. Then at most $1.5em$ main iterations need to be conducted to increase J_{max} and at most $1.5kem$ main iterations are needed when J_{max} reaches to k. Therefore, the maximal number of the criterion f_1 computing is $0.75pop \cdot kem$ to increase J_{max} to k. □

According to the Theorem 1 in [Qian et al. 2015], the number of the criterion f_1 computing is increased exponentially in POSS, which could be huge when the endmember number is large. However, it is increased linearly in MOSU. Since that the major computational load exists in the calculation of reconstruction error, MOSU is more suitable for further application.

4. Experiments

In this section, three synthetic and one real experiments are conducted to test the performance of MOSU. MOSU is compared with some state-of-the-art algorithms: two well-known convex relaxation methods (SUnSAL [Bioucas-Dias et al. 2010] and SUnSAL-TV [Iordache et al. 2012]) and three GAs (SMP [Shi et al. 2014], RSFoBa-Inf and RSFoBa-2 [Tang et al. 2014]). MOSU and POSS can always find a theoretically guaranteed approximation solution within the maximal iteration number, so we only compare their processing time. All the experiments are based on Chapter 1 of the United States Geological Survey (USGS)[1] digital spectral library (splib06a). The reflectance values of 498 materials are measured for 224 uniformly distributed spectral bands in the interval of $0.4 \sim 2.5\mu m$ ($A \in R^{224\times498}$). Details about the USGS library settings are described in the following experiments. Section 4.1 gives the performance evaluation criterions of experimental results. Section 4.2 describes the experiment environment and parameter setting

[1]Available online: http://speclab.cr.usgs.gov/spectral-lib.html

for all the involved algorithms. Section 4.3, 4.4 and 4.5 demonstrate the performance of MOSU on three synthetic data sets. Finally in Section 4.6, experiments on real data set are conducted and analyzed.

4.1. Performance evaluation criterion

To evaluate the abundance estimations for synthetic images, the root-mean-square error (RMSE) for each endmember and the signal to reconstruction error (SRE) for each pixel are used. RMSE for the ith endmember is described as below:

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_{i,j} - \hat{x}_{i,j})^2} \tag{30}$$

where $x_{i,j}$ represents the real abundance fraction for the ith endmember in the jth pixel, $\hat{x}_{i,j}$ represents the corresponding estimated value, n is the number of pixels in the synthetic image. The RMSE for each algorithm is the average of all endmember RMSEs:

$$\mathrm{RMSE} = \sum_{i=1}^{k}\mathrm{RMSE}_i \tag{31}$$

where k is the number of extracted endmembers and a smaller RMSE represents a better performance.

SRE provides a better measurement on the relationship between the power of error and signal as follows:

$$\mathrm{SRE} = 10log_{10}\left(\frac{\mathrm{E}[\|\mathbf{X}\|_F^2]}{\mathrm{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2]}\right) \tag{32}$$

where $\mathbf{X}$ is the true abundance matrix and $\hat{\mathbf{X}}$ is the estimated abundance matrix. For the SRE criterion, a larger value means a more accurate estimation.

4.2. Experiment environment and parameter setting

All the following experiments are executed on a computer with 3.40GHz Intel core, i7 CPU and 32GB RAM. The setting of some main parameters in MOSU are described here: Firstly, the sparsity k should be known in

advance. In this paper, we use HySime [Bioucas-Dias et al. 2008] algorithm to obtain this parameter automatically. Secondly, the *gen* in Algorithm 1 is set experientially based on the maximal number of main iterations in *Theorem 3.1*. Thirdly, the randomly flipping probability of each bit is set as $1/m$. According to Markov random process, the endmember extraction result would not be affected by $x/m, x = 1, 2, ...$ Lastly, in order to test the effect of population size in MOSU, we set it as 20 and 40 to compare their convergence speed in the following experiments. But all the other listed synthetic and real experiments are conducted with a 20 population size.

4.3. Experiments with synthetic data 1

In the first synthetic data experiments, 240 spectral vectors are randomly selected from the USGA library. Different signal-to-noise ratios (SNR) of noise are considered. The computation of SNR is:

$$\text{SNR} = 10log_{10}\left(\frac{\|\mathbf{Y}\|_F^2}{\|\mathbf{N}\|_F^2}\right) \tag{33}$$

where, a small SNR indicates a strong noise.

Considering the physical property of hyperspectral images that the noise is usually correlated, correlated noise is added to synthetic image data sets in this paper. The noise SNR is set as 20dB, 30dB and 40dB respectively.

Different endmember numbers are considered here, varying from 3 to 10. In each case, the corresponding synthetic image has the same image size of 64×64. The abundance fractions are generated based on the same operation in [Shi et al. 2014], where the values are forced to be smaller than 0.7 to avoid pure pixels. In this case, the active endmembers will be more difficult to be extracted.

Table 1 shows the SRE and RMSE results obtained by different algorithms. All the methods display well results when SNR is high. However, in low SNR data, the results of most methods present significant decline, while the proposed method decreases less. Specially, when the number of endmember is large, the superiority of MOSU is more obvious. Although the results of some greedy algorithm-based methods (SMP, RSFoBa-inf and RSFoBa-2) are close to that of MOSU when the number of endmember is small, superiority of the MOSU could be clearly observed with the increase of endmember number. Compared with convex relaxation-based methods (SUnSAL and SUnSAL-TV), the MOSU achieves better performance in nearly all the situation. Overall, MOSU performs better than other methods,

although the noise becomes gradually stronger and the endmember number becomes larger. The reason is that MOSU can extract optimal endmembers with little redundant endmember that is inactive in the image and the corresponding material does not exist in the image. Fortunately, the estimation of k is usually exactly in accordance with the truth in synthetic data cases, which results in no redundant endmembers for MOSU.

The mainly difference between MOSU and the greedy based SMP and RSFoBa (RSFoBa-Inf and RSFoBa-2) is the endmembers extracting. Unfortunately, SMP, RSFoBa-Inf and RSFoBa-2 always encounter the problem of missing active endmembers or extracting many redundant endmembers. The problem becomes more serious with the increase of endmember number and the decrease of SNR, especially in low SNR cases. However, MOSU can perform well on extracting endmembers and the results are always in accordance with the truth in synthetic experiments.

In order to further verify the performance of MOSU, the k top-ranking endmembers are recorded based on their residual values for SMP, RSFoBa-Inf and RSFoBa-2. Take the case of $k = 5$ as example, their top-five endmembers with relatively smaller residual values are selected respectively. We find that the top-five endmembers of all the three GAs are not in accordance with the truth. Therefore, it can be conclude that, SMP, RSFoBa-Inf and RSFoBa-2 cannot extract right endmembers although the active endmember number is estimated in advance. It can also be summarized that the bi-objective optimization based sparse unmixing strategy in MOSU is effective.

In addition, Fig. 3 shows the true abundance maps and the abundance maps estimated by all involved algorithms. The abundance maps are implemented with 30dB correlated noise when the endmember number is eight. Each row of Fig. 3 displays the abundance maps estimated by different algorithm for an endmember, and each column shows the results of an algorithm for different endmembers. It can be observed that all the methods perform well in most cases. However, compared with other methods, the results of MOSU have less noise. Moreover, from Fig. 3(f) we can see that the MOSU perform better in preserving the outline information. In general, the MOSU achieves the closest results to the truth.

The processing time of MOSU and POSS on synthetic data 1 with 30dB correlated noise and 3$\sim$10 endmembers is listed in Table 2. It is the average time after running each algorithm ten times. To analysis the effect of population number in MOSU, we set it as 20, 40 respectively and denote the corresponding algorithms as MOSU20, MOSU40. From Table 2, it can be

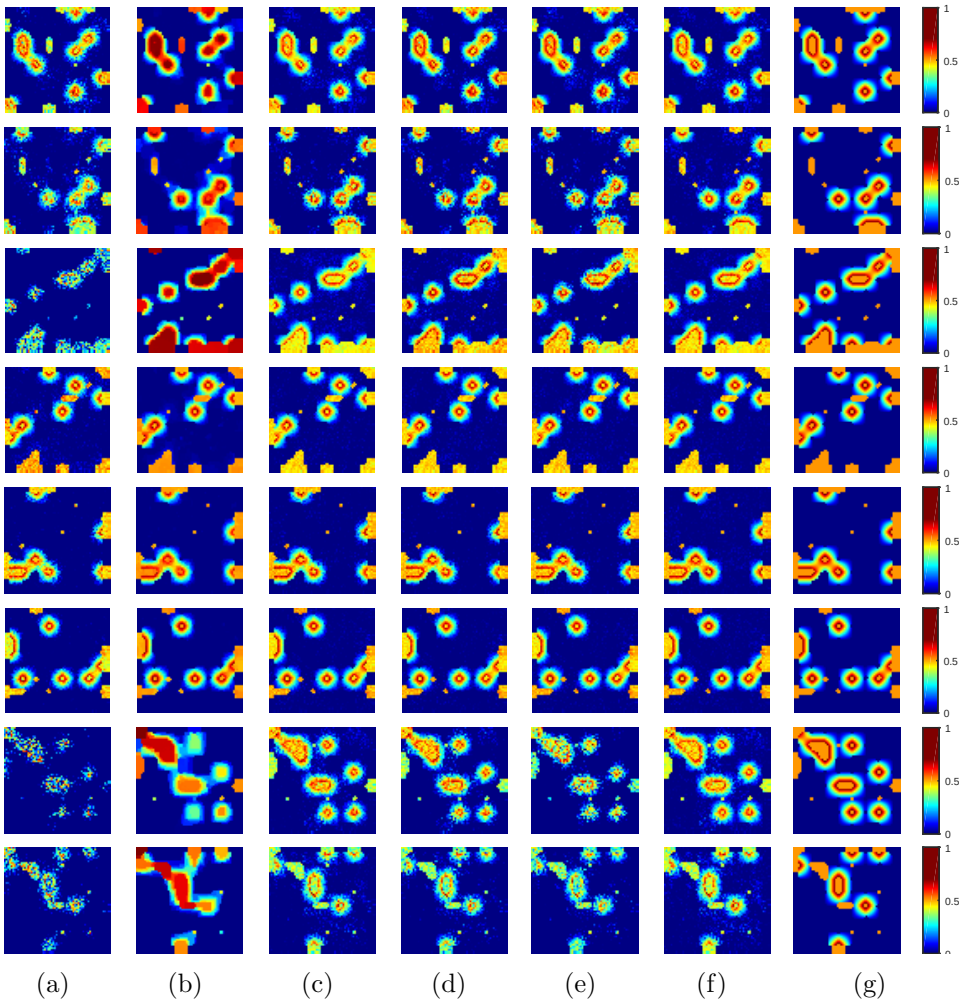

Fig. 3: Comparison of abundance maps on synthetic data 1 with 30dB correlated noise when the endmember number is eight. From top row to bottom row are are the maps corresponding to endmembers from 1 to 8. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-Inf, RSFoBa-2, MOSU and the truth respectively.

observed that MOSU20 converges faster than POSS in most cases. MOSU40 is usually slower than MOSU20, even slower than POSS especially when the endmember number is large. Therefore, MOSU20 with a 20 population size is more suitable for sparse unmixing.

Table 1: SRE and RMSE results for synthetic data 1 obtained by different algorithms.

SRE for Synthetic Data 1 obtained by different algorithms.							
		SUnSAL	SUnSAL-TV	SMP	RSFoBa-inf	RSFoBa-2	MOSU
20dB	3	2.8076	3.5098	15.3646	15.3646	11.4518	15.3646
	4	2.9599	4.2180	11.0828	13.5643	10.9240	13.5643
	5	4.5366	6.0651	11.5581	12.6789	5.6186	12.6789
	6	3.9672	5.2114	11.7837	7.9009	5.6513	11.7837
	7	2.6258	3.9804	4.7634	2.8868	2.0382	11.0265
	8	1.9617	3.7398	4.5147	3.9948	3.9948	9.3688
	9	2.9957	4.4994	3.3664	3.1914	-0.6123	9.0067
	10	2.8201	4.2564	3.9317	0.5114	0.0231	9.0858
30dB	3	8.2087	14.2316	25.0731	25.0731	21.7841	25.0731
	4	7.9690	14.7605	23.2740	23.2740	20.5890	23.2740
	5	9.4310	15.2156	20.6604	22.2056	16.6541	22.2056
	6	8.6034	13.9011	20.0220	19.6872	14.2256	21.0834
	7	5.8326	11.9027	15.6379	11.5078	6.0484	20.2018
	8	5.4090	11.2326	15.3641	12.9509	8.5018	17.8117
	9	6.3004	12.0434	11.2738	7.9266	7.5480	17.7749
	10	6.0235	11.7989	6.5884	3.0025	3.0679	17.9527
40dB	3	17.7541	21.7765	35.0535	35.0535	35.0535	35.0535
	4	17.4207	21.6018	33.0989	33.0989	30.5111	33.0989
	5	18.1889	22.5929	30.3167	32.0162	24.7172	32.0162
	6	16.7907	21.0164	29.9143	30.9952	24.9827	30.9952
	7	13.1550	19.4718	25.2325	27.4613	16.6172	30.0172
	8	12.7155	18.4439	24.4671	26.3354	14.2802	27.7430
	9	13.0964	19.1598	11.2738	13.7984	13.7984	27.4860
	10	13.1162	18.7945	11.1832	4.4256	4.3458	27.5013
RMSE for synthetic data 1 obtained by different algorithms.							
		SUnSAL	SUnSAL-TV	SMP	RSFoBa-inf	RSFoBa-2	MOSU
20dB	3	0.2302	0.2410	0.0626	0.0626	0.0865	0.0626
	4	0.2015	0.1561	0.1028	0.0704	0.0888	0.0704
	5	0.1580	0.1220	0.0851	0.0675	0.1116	0.0675
	6	0.1434	0.1133	0.0697	0.0806	0.1018	0.0697
	7	0.2794	0.2481	0.2397	0.2283	0.2475	0.1927
	8	0.1541	0.1182	0.1041	0.1299	0.1299	0.0757
	9	0.1323	0.1075	0.1290	0.1293	0.1369	0.0759
	10	0.1244	0.0995	0.1067	0.1039	0.1057	0.0698
30dB	3	0.1212	0.0728	0.0205	0.0205	0.0285	0.0205
	4	0.1048	0.0546	0.0296	0.0231	0.0296	0.0231
	5	0.0824	0.0457	0.0304	0.0225	0.0355	0.0225
	6	0.0750	0.0437	0.0292	0.0252	0.0406	0.0238
	7	0.0972	0.0519	0.0534	0.0424	0.0452	0.0237
	8	0.0952	0.0524	0.0336	0.0424	0.0552	0.0280
	9	0.0809	0.0452	0.0459	0.0511	0.0530	0.0274
	10	0.0775	0.0422	0.0692	0.0672	0.0657	0.0249
40dB	3	0.0403	0.0277	0.0065	0.0065	0.0065	0.0065
	4	0.0352	0.0233	0.0074	0.0074	0.0094	0.0074
	5	0.0297	0.0188	0.0100	0.0073	0.0132	0.0073
	6	0.0284	0.0180	0.0093	0.0076	0.0124	0.0076
	7	0.0399	0.0208	0.0167	0.0092	0.0167	0.0076
	8	0.0394	0.0223	0.0166	0.0096	0.0223	0.0090
	9	0.0348	0.0194	0.0459	0.0256	0.0256	0.0089
	10	0.0328	0.0182	0.0286	0.0391	0.0396	0.0082

Table 2: The computational time of MOSUs and POSS (sec).

endmember number	MOSU20	MOSU40	POSS
3	26.45	67.89	75.97
4	181.84	148.90	162.96
5	193.15	235.86	262.63
6	234.09	289.08	242.48
7	357.29	431.77	408.55
8	348.20	486.13	389.30
9	551.98	476.47	695.22
10	536.11	666.64	617.00

4.4. Experiments with synthetic data 2

In the second synthetic data experiments, the spectral library also includes 240 signatures. Among them, five similar spectral signatures are selected specially from the USGA library and the others are selected from the library used in synthetic data 1. The first five signatures are Actinolite HS116.3B, Actinolite HS22.3B, Actinolite HS315.4B, Actinolite NMNH80714 and Actinolite NMNHR16485. Unmixing problem is more difficult in this case. For example, Actinolite HS22.3B is likely to be missed when Actinolite HS116.3B is selected because they are too similar. It can also happen among the rest spectral signatures. Their spectra are shown in Fig. 4. The synthetic image here is generated based on the same operation in the above synthetic experiments. The size of all the generated images is also 64×64. The noise SNR here are 20dB, 30dB and 40dB respectively and the endmember number is varying from 3 to 10.

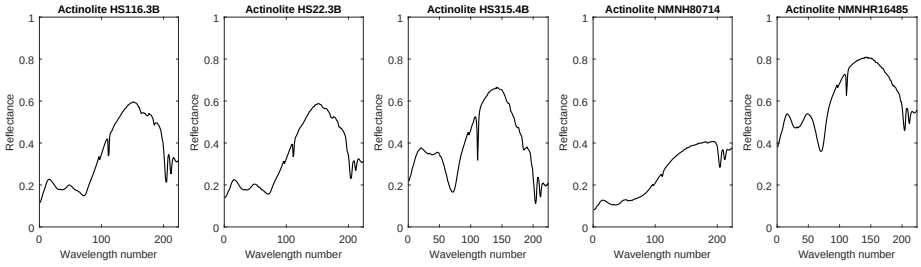

Fig. 4: Spectra of five selected highly similar materials in synthetic data 2.

Table 3 shows results obtained by different algorithms. It can be observed that MOSU also performs best. Performance of the algorithms are tend to be worse than that on synthetic data 1. In Table 3, the advantage of the MOSU is also obvious, especially when the endmenber number is large. Furthermore, we observe that different endmember number has little influence on the results of MOSU. This phenomenon indicates that the MOSU is not very sensitive to the variation of endmember number. In addition, Fig. 5 shows the true abundance maps and the abundance maps estimated by all the algorithms on synthetic data 2. The abundance maps is also implemented with 30dB correlated noise when the endmember number is eight. Similar to the results of Fig. 3, in Fig. 5, we can see that the abundance maps of MOSU show better outline preserving capacity. The three greedy algorithm-based methods present close performance, and their results are closer to the truth, compared with SUnSAL and SUnSAL-TV. However, MOSU outperforms all of them in most cases.

4.5. Experiments with synthetic data 3

The third synthetic dataset is provided by [Iordache et al. 2012] using five randomly selected spectral signatures from the library in Section 4.3. Synthetic data 3 contains 75×75 pixels, whose background pixels are mixed by the five selected pixels with the fractional abundance values 0.1149, 0.0741, 0.2003, 0.2055, and 0.4051, respectively.

Table 4 shows SRE and RMSE results on synthetic data 3 obtained by different algorithms. It can be observed that MOSU has the best SREs and RMSEs among the six algorithms. Fig. 6 grapes the abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-Inf, RSFoBa-2, MOSU and the truth with 30dB correlated noise. Each row of Fig. 6 are abundance maps obtained by an algorithm, and each column corresponds to the results of different algorithms for an endmember. We can see that the estimated abundance maps obtained by SUnSAL, SMP, RSFoBa-Inf and RSFoBa-2 have more noise, especially in the abundance maps of endmember 5. By comparison, the estimated results of SUnSAL-TV and MOSU are closer to the truth. Although the unmixing results of SUnSAL-TV have less noise than MOSU, more patches in its abundance maps are lost in endmember 1 and endmember 3. Overall, MOSU can obtain a relatively better results compared with the others.

Table 3: SRE and RMSE results for synthetic data 2 obtained by different algorithms.

SRE for Synthetic Data 2 obtained by different algorithms							
		SUnSAL	SUnSAL-TV	SMP	RSFoBa-inf	RSFoBa-2	MOSU
20dB	3	1.4935	4.7485	1.7569	1.7569	8.7495	9.1863
	4	0.8008	3.8189	0.2944	-0.1087	1.2529	9.6566
	5	1.0132	3.5134	0.5111	0.5111	0.6152	8.4236
	6	1.4762	3.5906	1.2019	1.5000	1.9012	8.7361
	7	0.9016	2.9262	1.6817	1.6817	1.4331	9.1976
	8	0.8754	3.2240	2.8397	4.4345	2.1443	8.8730
	9	0.7792	2.9297	0.1262	1.5539	2.8646	8.0299
	10	1.3463	3.2135	0.8055	0.8055	2.2762	7.8837
30dB	3	4.5008	6.3980	3.8972	6.6002	3.2224	18.7641
	4	4.2866	5.6305	1.4631	2.0494	4.1208	19.2000
	5	3.5375	5.3574	2.0818	1.9945	3.9870	17.5236
	6	3.8292	5.8552	2.6497	3.4193	2.6282	16.9777
	7	3.4025	5.9792	2.6299	3.1775	3.1049	17.8688
	8	3.5173	6.4997	5.1362	4.9291	5.9503	17.5240
	9	2.7543	5.4745	1.4363	4.2383	4.1972	16.1698
	10	3.4184	5.9015	1.9468	5.0523	4.2713	16.1400
40dB	3	10.6950	6.7210	9.0856	28.5543	3.3601	28.5543
	4	9.4186	5.8792	2.6919	9.0453	25.9514	29.0152
	5	8.2928	5.6425	4.4459	4.4549	5.8321	27.2800
	6	7.9923	6.3070	4.8654	4.9972	5.8857	26.7131
	7	7.8238	6.7670	5.0718	8.2771	9.6493	27.5575
	8	7.9719	7.7274	5.0716	5.8663	5.8666	26.7139
	9	7.0023	7.2006	2.5629	7.8910	9.2511	25.8656
	10	7.5349	8.0748	2.7838	8.6649	8.6062	25.4837
RMSE for synthetic data 2 obtained by different algorithms							
		SUnSAL	SUnSAL-TV	SMP	RSFoBa-inf	RSFoBa-2	MOSU
20dB	3	0.2618	0.1899	0.2391	0.2391	0.1239	0.1193
	4	0.2555	0.1933	0.2634	0.3239	0.2838	0.0979
	5	0.2254	0.1768	0.2309	0.2309	0.2265	0.1035
	6	0.1896	0.1525	0.1796	0.1731	0.1695	0.0864
	7	0.1898	0.1548	0.1561	0.1561	0.1627	0.0776
	8	0.1780	0.1350	0.1282	0.1090	0.1419	0.0799
	9	0.1700	0.1352	0.1516	0.1333	0.1189	0.0803
	10	0.1428	0.1168	0.1509	0.1509	0.1210	0.0762
30dB	3	0.1726	0.1342	0.1644	0.0966	0.1958	0.0394
	4	0.1612	0.1340	0.2224	0.1961	0.1857	0.0324
	5	0.1559	0.1248	0.1818	0.1821	0.1429	0.0360
	6	0.1329	0.1036	0.1433	0.1236	0.1450	0.0329
	7	0.1351	0.0988	0.1129	0.1207	0.1285	0.0283
	8	0.1233	0.0852	0.0839	0.0897	0.0807	0.0294
	9	0.1285	0.0928	0.1459	0.0883	0.0889	0.0308
	10	0.1068	0.0803	0.1276	0.0757	0.0867	0.0290
40dB	3	0.0795	0.1131	0.0861	0.0128	0.1810	0.0128
	4	0.0824	0.1160	0.1861	0.0526	0.0157	0.0104
	5	0.0838	0.1082	0.1323	0.1323	0.1038	0.0117
	6	0.0751	0.0859	0.1096	0.1105	0.0957	0.0107
	7	0.0755	0.0778	0.0919	0.0612	0.0502	0.0093
	8	0.0681	0.0633	0.0636	0.0755	0.0755	0.0101
	9	0.0732	0.0636	0.1329	0.0491	0.0409	0.0101
	10	0.0620	0.0522	0.1000	0.0425	0.0428	0.0099

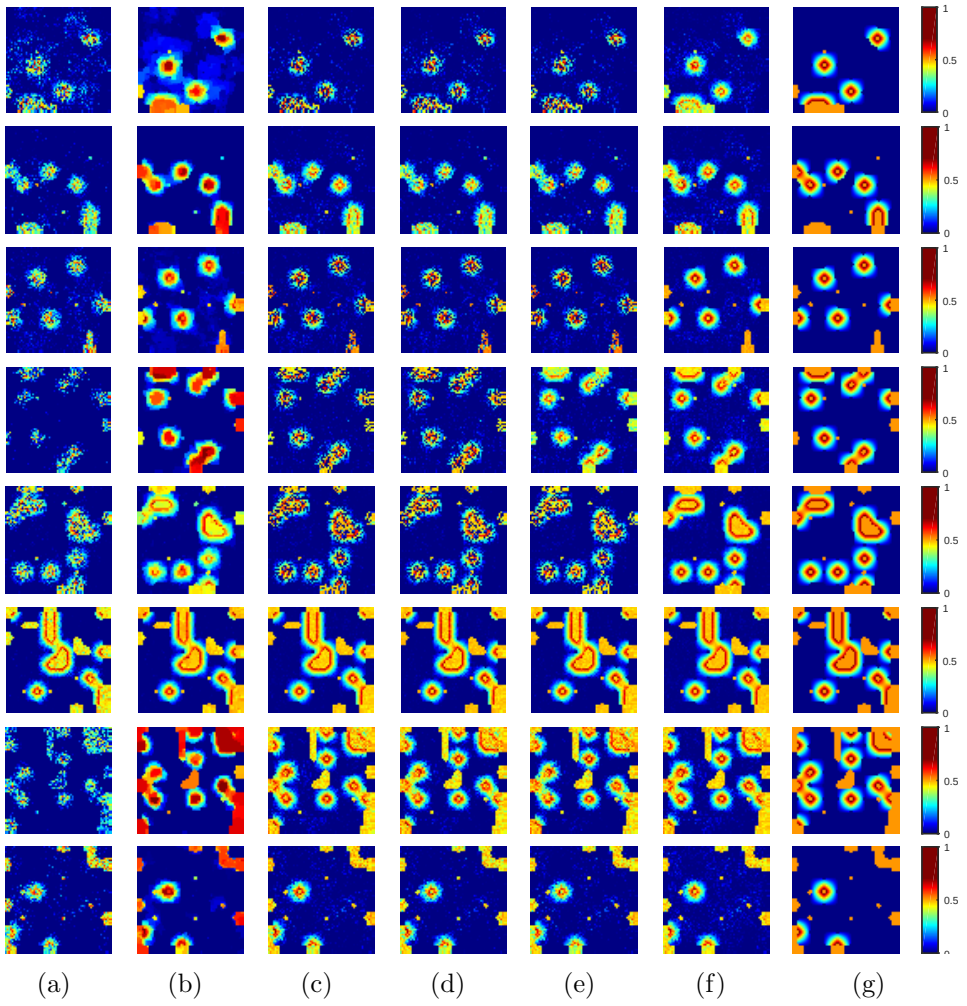

Fig. 5: Comparison of abundance maps on synthetic data 2 with 30dB correlated noise when the endmember number is eight. From top row to bottom row are are the maps corresponding to endmembers from 1 to 8. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-Inf, RSFoBa-2, MOSU and the truth respectively.

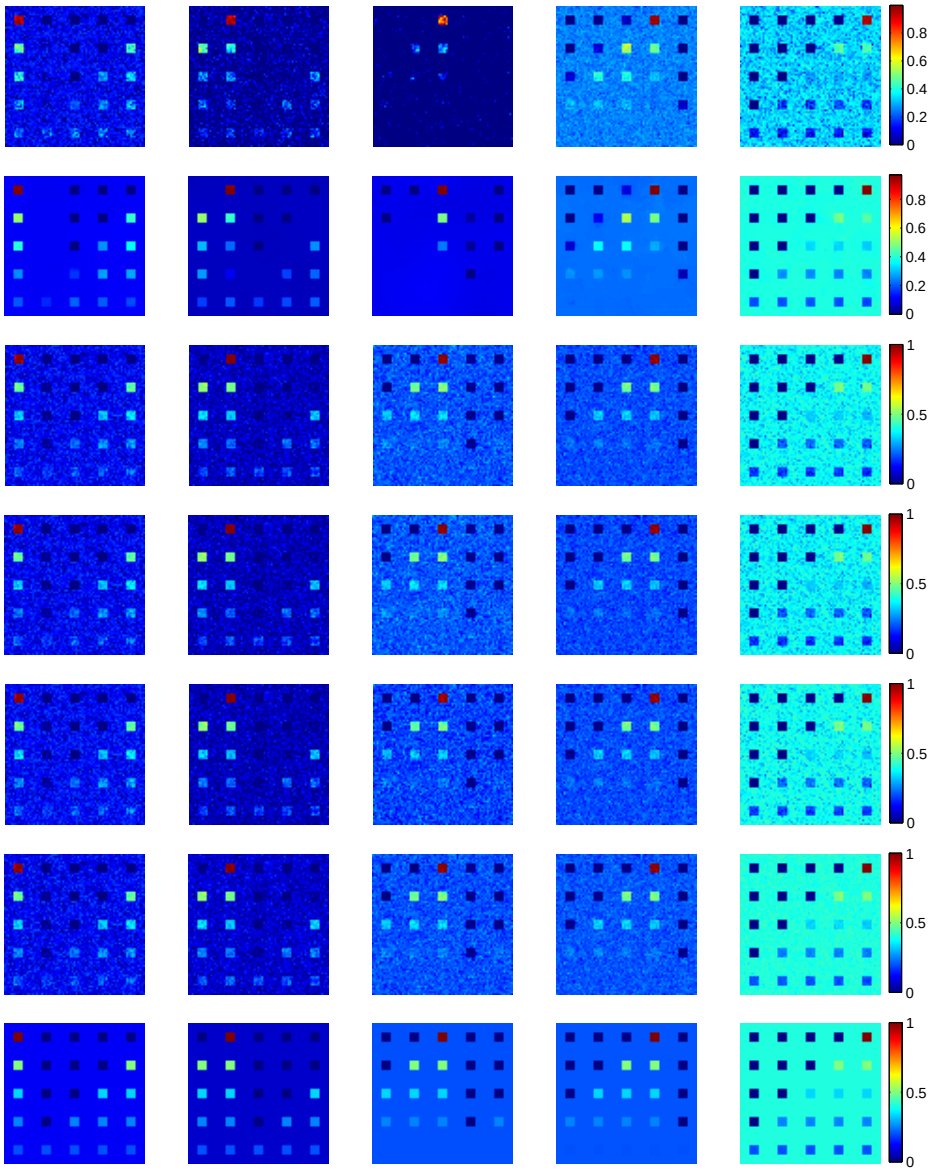

Fig. 6: Comparison of abundance maps on synthetic data 3 with 30dB correlated noise. From top row to bottom row are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-Inf, RSFoBa-2, MOSU and the truth respectively. From left column to right column are the maps corresponding to endmembers from 1 to 5.

Table 4: SRE and RMSE results for synthetic data 3 obtained by different algorithms.

SRE for Synthetic Data 3 obtained by different algorithms						
SNR	SUnSAL	SUnSAL-TV	SMP	RSFoBa-inf	RSFoBa-2	MOSU
20dB	2.8740	6.1687	5.2260	5.5423	4.2532	9.2787
30dB	6.1289	10.4846	13.5314	12.6625	11.7148	16.7099
40dB	11.0483	20.8034	23.0685	22.8953	22.7590	26.4514
RMSE for Synthetic Data 3 obtained by different algorithms						
SNR	SUnSAL	SUnSAL-TV	SMP	RSFoBa-inf	RSFoBa-2	MOSU
20dB	0.1405	0.0845	0.1001	0.0952	0.1004	0.0798
30dB	0.0925	0.0463	0.0413	0.0439	0.0424	0.0334
40dB	0.0508	0.0179	0.0138	0.0140	0.0141	0.0108

4.6. Experiments with real data

Many real hyperspectral datasets have provided a realistic opportunity to explore the methodology of hyperspectral data analysis. The well-known AVIRIS Cuprite dataset with 224 spectral bands is used in this paper. A hyperspectral image used here is a 204×151 subscene of the Cuprite dataset[2] with 188 spectral bands collected in 1997, where some water absorption and low SNR bands are removed. A Tricorder map[3] produced by a USGS in 1995 is shown in Fig. 7 to further illustrate the image. The mineral distribution maps of Alunite+Muscovite, Alunite, Hematite and Montmorillonite produced by Tricorder software for the 204×151 AVIRIS Cuprite scene is shown in Fig. 8. Tricorder map labels each pixel pure for one special material. But an abundance fraction is used to represent the degree of each mineral presented in the pixel in the unmixing result for each pixel. Therefore, the mineral map is only a qualitative assessment of the algorithm performance and cannot be used as a direct comparison between algorithms. The spectral library used here is different from that in synthetic experiments. The USGS spectral library consists of 498 minerals with the above corresponding bands removed.

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

Considering that there always exist a large number of materials in real hyperspectral image, the Cuprite image is divided into several blocks to get better performances. The block size for MOSU is set to 12. HySime is used to estimate endmember number in the image dataset and the result is 14. Considering the complexity in real-world images, we just take the HySime result as a reference and set the whole endmember number larger than 14 to ensure little endmember missed. Thus, we set k in each block as 2, 3 and term the resulting algorithms as MOSU2, MOSU3, respectively. All the five comparison methods used above are also discussed in this real dataset. Abundance maps of the above four minerals (Alunite+Muscovite, Alunite, Hematite and Montmorillonite) estimated by the seven algorithms are shown in Fig. 9 and Fig. 10. We can see that the abundances estimated by MOSU2 are comparable or higher in the regions assigned to the corresponding minerals, compared to the other considered algorithms. SMP and RSFoBa-2 cannot find the Alunite mineral. SUnSAL, SUnSAL and RSFoBa-Inf cannot give the abundance of Alunite a good estimation. Only MOSU3 and MOSU2 give the Montmorillonite distribution clear descriptions. Generally, MOSU2 is slightly better than MOSU3, cause there exist less redundant endmembers in it. Thus, it can be concluded that the proposed multi-objective optimization based MOSU unmixing method is valid in real-world hyperspecral image unmixing.

5. Conclusion

To solve the sparse unmixing problem, in this paper, a multi-objective optimization based sparse unmixing algorithm (MOSU) for hyperspectral image is proposed. MOSU transforms sparse unmixing to a bi-objective optimization problem. The reconstruction error and the sparsity of endmember are the two objectives. In this case, the l_0 norm based sparse unmixing can be handled directly without relaxation. The proposed unmixing method is inspired by a Pareto optimization based POSS method that can be used to solve subset selection problems. We extend POSS to a population-based algorithm to improve its computation efficiency and demonstrate the effectivity theoretically. Experiments on the synthetic datasets and the real-world dataset verify the good performance of MOSU. In conclusion, MOSU have good performance in solving the sparse unmixing problem, and we will focus on exploring more effective methods to solve this problem.

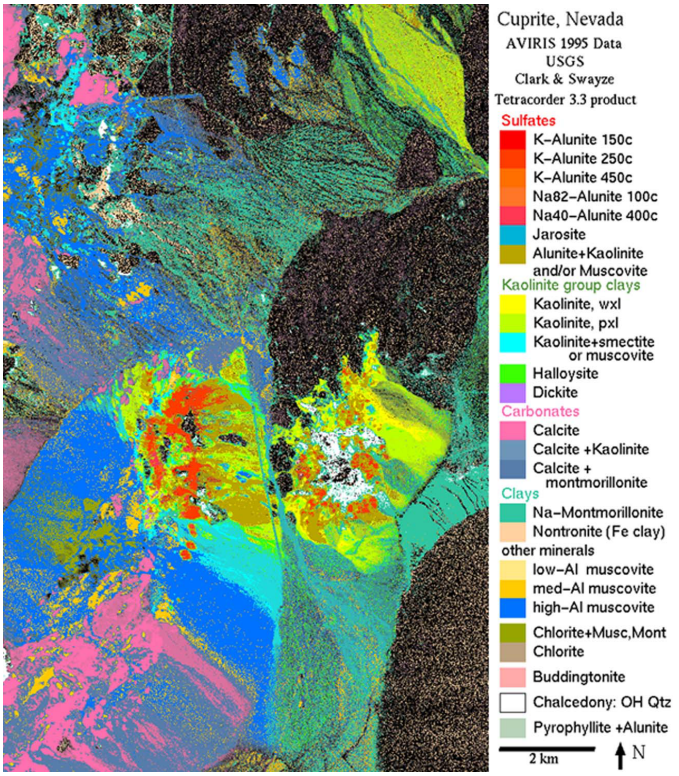

Fig. 7: USGS map showing the location of different minerals in the Cuprite mining district in NV.

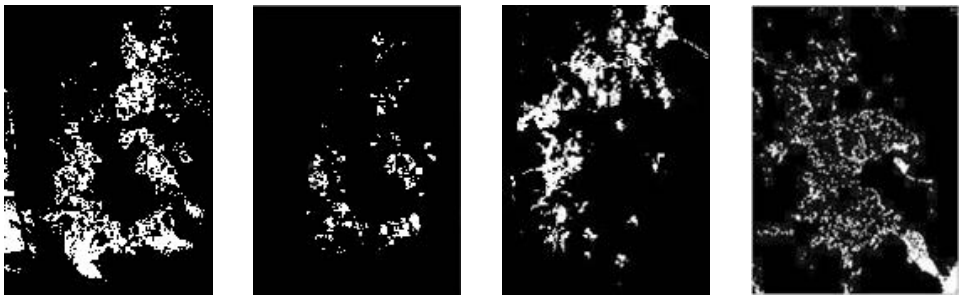

Fig. 8: The classification maps produced by the USGS Tricorder algorithm for the 204×151 AVIRIS Cuprite scene.

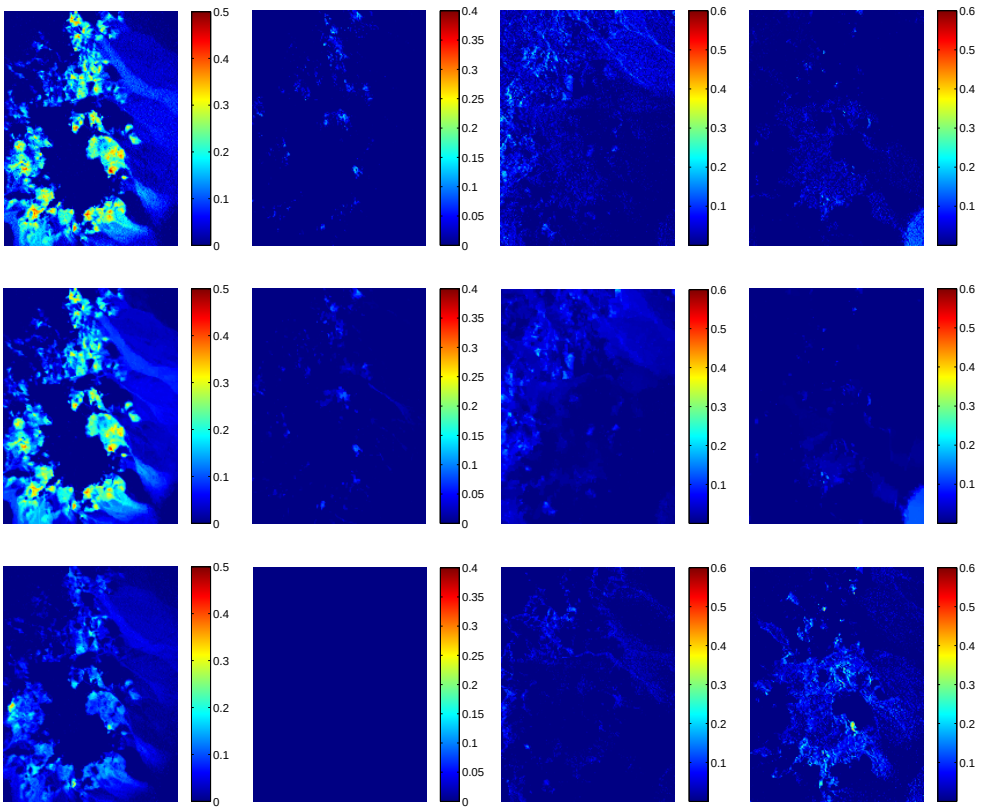

Fig. 9: From top row to bottom row are abundance maps estimated by SUnSAL, SUnSAL-TV, SMP on the 204×151 subscene of the Cuprite dataset. From left column to right column are the maps corresponding to Alunite+Muscovite, Alunite, Hematite and Montmorillonite, respectively.

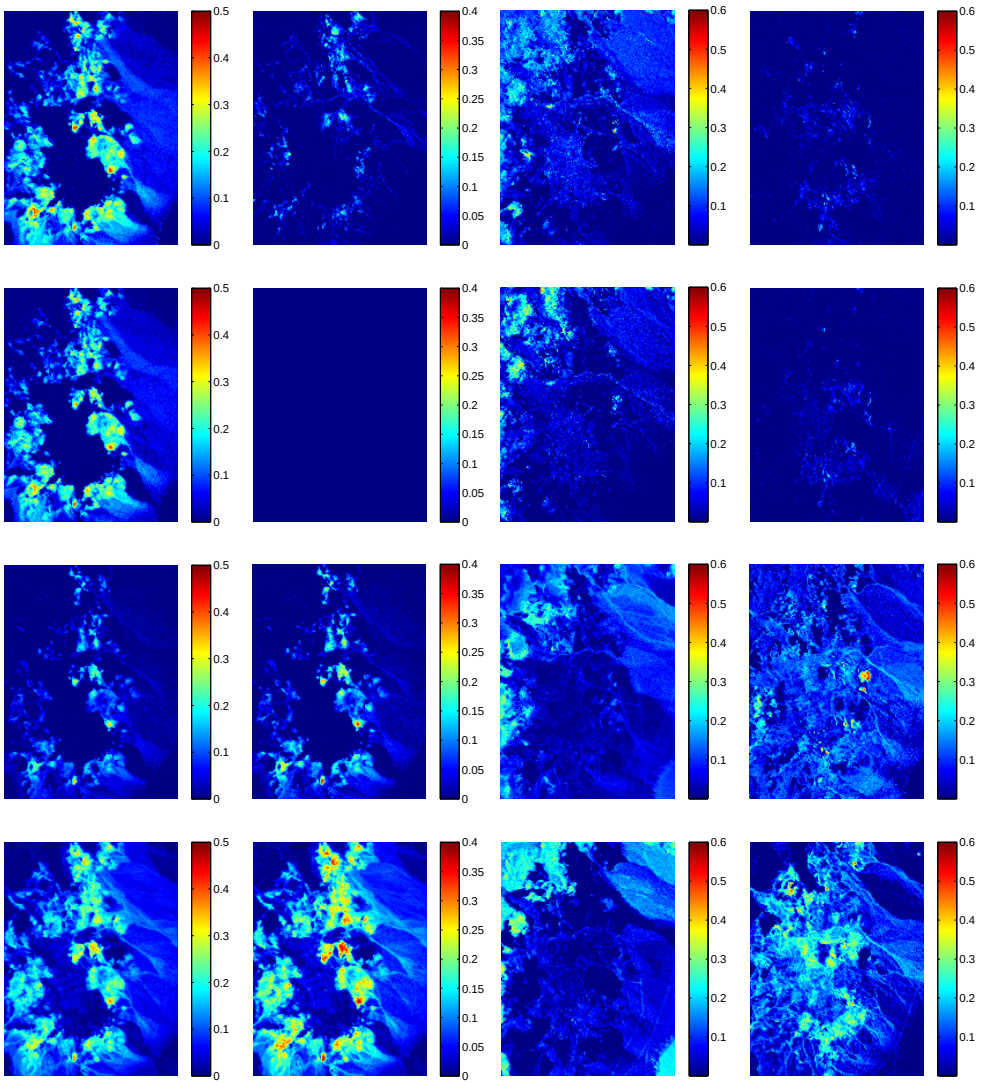

Fig. 10: From top row to bottom row are abundance maps estimated by RSFoBa-Inf, RSFoBa-2, MOSU3 and MOSU2 on the 204×151 subscene of the Cuprite dataset. From left column to right column are the maps corresponding to Alunite+Muscovite, Alunite, Hematite and Montmorillonite, respectively.

References

[Bioucas-Dias et al. 2012] Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J., 2012. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. IEEE J. Sel. Topics Appl. Earth Observ. Rem. Sens. 5(2), 354–379.

[Green et al. 1998] Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M., Olah, M.R., Williams, O., 1998. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). Rem. Sens. Environ. 65(3), 227–248.

[Keshava et al. 2002] Keshava, N., Mustard, J.F., 2002. Spectral unmixing. IEEE Signal Process. Mag. 19(1), 44–57.

[Shippert. 2004] Shippert, P., 2004. Why use hyperspectral imagery? Photogramm. Eng. Rem. Sens. 70(4), 377–396.

[Landgrebe. 2002] Landgrebe, D., 2002. Hyperspectral image data analysis, IEEE Signal Process. Mag. 19(1), 17–28.

[Bioucas-Dias et al. 2013] Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci. Rem. Sens. Mag. 1(2), 6–36.

[Hu et al. 1999] Hu, Y., Lee, H., Scarpace, F., 1999. Optimal linear spectral unmixing. IEEE Trans. Geosci. Rem. Sens. 37(1), 639–644.

[Petrou et al. 1999] Petrou, M., Foschi, P.G., 1999. Confidence in linear spectral unmixing of single pixels. IEEE Trans. Geosci. Rem. Sens. vol. 37(1), 624–626.

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

[Zhong et al. 2016] Zhong, Y.F., Wang, X.Y., Zhao, L., Feng, R.Y., Zhang, L.P., Xu, Y.Y., 2016. Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery. ISPRS J. Photogram. Rem. Sens. 119, 49–63.

[Parra et al. 1999] Parra, L., Spence, C., Sajda, P., Ziehe, A., Muller, K.R., 1999. Unmixing hyperspectral data. in Proc. Adv. Neural Inf. Process. Syst. 942–948.

[Adams et al. 1986] Adams, J.B., Smith, M.O., Johnson, P.E., 1986. Spectralmixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site. J. Geophys. Res. 91, 8098–8112.

[Ma et al. 2014] Ma, W.K., Bioucas-Dias, J.M., Chan, T.H., Gillis, N., Gader, P., Plaza, A., Ambikapathi, A., Chi, C.Y., 2014. A signal processing perspective on hyperspectral unmixing: Insights from Remote Sensing. IEEE Signal Process. Mag. 31(1), 67–81.

[Boardman. 1993] Boardman, J., 1993. Automating spectral unmixing of AVIRIS data using convex geometry concepts. in Proc. Summaries 4th Annu. JPL Airborne Geoscience Workshop. 1(JPL Pub. 93-26), 11–14.

[Winter. 2003] Winter, M.E., 2003. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. in Proc. SPIE Imaging Spectrometry V. 3753, 266–275.

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

[Nascimento et al. 2005] Nascimento, J., Bioucas-Dias, J., 2005. Does independent component analysis play a role in unmixing hyperspectral data? IEEE Trans. Geosci. Rem. Sens. 43(1), 175–187.

[Qian et al. 2011] Qian, Y., Jia, S., Zhou, J., Robles-Kelly, A., 2011. Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization. IEEE Trans. Geosci. Rem. Sens. 49(11), 4282–4297.

[Lu et al. 2013] Lu, X., Wu, H., Yuan, Y., Yan, P., Li, X., 2013. Manifold regularized sparse NMF for hyperspectral unmixing. IEEE Trans. Geosci. Rem. Sens. 51(5), 2815–2826.

[Zhu et al. 2014] Zhu, F.Y., Wang, Y., Xiang, S.M., Fan, B., Pan, C.H., 2014. Structured sparse method for hyperspectral unmixing. ISPRS J. Photogram. Rem. Sens. 88, 101–118.

[Iordache et al. 2011] Iordache, M.D., Bioucas-Dias, J.M., Plaza, A., 2011. Sparse unmixing of hyperspectral data. IEEE Trans. Geosci. Rem. Sens. 49(6), 2014–2039.

[Berman et al. 2004] Berman, M., Kiiveri, H., Lagerstrom, R., Ernst, A., Dunne, R., Huntington, J.F., 2004. ICE: A statistical approach to identifying endmembers in hyperspectral images. IEEE Trans. Geosci. Rem. Sens. 42(10), 2085–2095.

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

[Elad. 2010] Elad, M., 2010. Sparse and redundant representations. New York, NY, USA: Springer-Verlag.

[Gong et al. 2016] Gong, M.G., Li, H., Luo, E.H., Liu, J., Liu, J., 2016. A Multi-objective cooperative coevolutionary algorithm for hyperspectral sparse unmixing. IEEE Trans. Evol. Comput. 99(PP).

[Eldar et al. 2010] Eldar, Y.C., Rauhut, H., 2010. Average case analysis of multichannel sparse recovery using convex relaxation. IEEE Trans. Inf. Theory. 56(1), 505–519.

[Tropp. 2006] Tropp, J.A., 2006. Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. EURASIP J. Signal Process. 86(3), 589–602.

[Tropp. 2004] Tropp, J.A., 2004. Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Inf. Theory. 50(10), 2231–2242.

[Tropp et al. 2006] Tropp, J.A., Gilbert, A.C., Strauss, M.J., 2006. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit, EURASIP J. Signal Process. 86(3), 572–588.

[Esser. 2009] Esser, E., 2009. Applications of lagrangian-based alternating direction methods and connections to split bregman. Univ. California Los Angeles, Los Angeles, CA, USA, CAM Rep. 9–31.

[Yang et al. 2010] Yang, J., Zhang, Y., 2010. Alternating direction algorithms for l1-problems in compressive sensing, Rice Univ., Houston, TX, USA, Tech. Rep. CAAM TR09–37.

[Iordache et al. 2014] Iordache, M.D., Bioucas-Dias, J.M., Plaza, A., 2014. Collaborative sparse regression for hyperspectral unmixing. IEEE Trans. Geosci. Rem. Sens. 52(1), 341–354.

[Iordache et al. 2012] Iordache, M.D., Bioucas-Dias, J.M., Plaza, A., 2012. Total variation spatial regularization for sparse hyperspectral unmixing. IEEE Trans. Geosci. Rem. Sens. 50(11), 4484–4502.

[Feng et al. 2014] Feng, R.Y., Zhong, Y.F., Zhang, L.P., 2014. Adaptive nonlocal Euclidean medians sparse unmixing for hyperspectral imagery, ISPRS J. Photogram. Rem. Sens. 97, 9–24.

[Feng et al. 2016] Feng, R.Y., Zhong, Y.F., Zhang, L.P., 2016. Adaptive spatial regularization sparse unmixing strategy based on joint MAP for hyperspectral remote sensing imagery, IEEE J. Sel. Topics Appl. Earth Observ. Rem. Sens. 9(12), 5791–5805.

[Feng et al. 2016] Feng, R.Y., Zhong, Y.F., Wu, Y.Y., He, D., Xu, X., Zhang, L.P., 2016. Nonlocal total variation subpixel mapping for hyperspectral remote sensing imagery, Rem. Sens. 8(3), 250.

[Tang et al. 2015] Tang, W., Shi, Z., Wu, Y., Zhang, C., 2015. Sparse unmixing of hyperspectral data using spectral a priori information. IEEE Trans. Geosci. Rem. Sens. 53(2), 770–783.

[Shi et al. 2014] Shi, Z.W., Tang, W., Duren, Z., Jiang., Z.G., 2014. Subspace matching pursuit for sparse unmixing of hyperspectral data. IEEE Trans. Geosci. Rem. Sens. 52(6), 3256–3274.

[Tang et al. 2014] Tang, W., Shi, Z.W., Wu, Y., 2014. Regularized simultaneous forward-backward greedy algorithm for sparse unmixing of hyperspectral data. IEEE Trans. Geosci. Rem. Sens. 52(9), 5271–5288.

[Themelis et al. 2012] Themelis, K.E., Rontogiannis, A.A., Koutroumbas, K.D., 2012. A Novel Hierarchical Bayesian Approach for Sparse Semi-Supervised Hyperspectral Unmixing. IEEE Trans. Signal Process. 60(2), 585–599.

[Deb et al. 2001] Deb, K., Kalyanmoy, D., 2001. Multi-objective optimization using evolutionary algorithms. Wiley and Sons. 2(3), 509.

[Sindhya et al. 2013] Sindhya, K., Miettinen, K., Deb., K., 2013. A hybrid framework for evolutionary multi-objective optimization. IEEE Trans. Evol. Comput. 17(4), 495–511.

[Qian et al. 2015] Qian, C., Yu, Y., Zhou, Z.H., 2015. Subset selection by pareto optimization. in Proc. Adv. Neural Inf. Process. Syst. 1765–1773.

[Deb et al. 2002] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist mul-tiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6(2), 182–197.

[Li et al. 2013] Li, Y.F., Pedroni, N., Zio, E., 2013. A memetic evolutionary multi-objective optimization method for environmental power unit commitment. IEEE Trans. Power Syst. 28(3), 2660–2669.

[Xu et al. 2014] Xu, Y., Dong, Z.Y., Meng, K., Yao, W.F., Zhang, R., Wong, K.P., 2014. Multi-objective dynamic VAR planning against short-term voltage instability using a decomposition-based evolutionary algorithm. IEEE Trans. Power Syst. 29(6), 2813–2822.

[Eckart et al. 1999] Eckart, Z., Lothar, T., 1999. Multi-objective evolutionary algorithms: A comparative case study and the strength Pareto approach. IEEE Trans. Evol. Comput. 3(4), 257–271.

[Siinivas et al. 1994] Siinivas, N., Deb, K., 1994. Multi-objective optimization using nondominated sorting in genetic algorithms. Evol. Comput. 2(3), 221–248.

[Das et al. 2011] Das, A., Kempe, D., 2011. Submodular meets Spectral: Greedy Algorithms for Subset Selection, Sparse Approximation and Dictionary Selection. Comput. Sci.

[Das et al. 2008] Das, A., Kempe, D., 2008. Algorithms for Subset Selection in Linear Regression. in Proc. STOC, 45–54.

[Bioucas-Dias et al. 2010] Bioucas-Dias, J.M., Figueiredo, M.,2010. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. in Proc. 2nd Workshop Hyperspectr. Image Signal Process. Evol. Rem. Sens. 1, 1–4.

[Bioucas-Dias et al. 2008] Bioucas-Dias, J.M., Nascimento, J.M.P., 2008. Hyperspectral subspace identification. IEEE Trans. Geosci. Rem. Sens. 46(8), 2435–2445.