

# A Classification-Based Model for Multi-Objective Hyperspectral Sparse Unmixing

Xia Xu<sup>1</sup>, Zhenwei Shi<sup>1</sup>, *Member, IEEE*, Bin Pan<sup>1</sup>, and Xuelong Li<sup>2</sup>, *Fellow, IEEE*

**Abstract**—Sparse unmixing has become a popular tool for hyperspectral imagery interpretation. It refers to finding the optimal subset of a spectral library to reconstruct the image data and further estimate the proportions of different materials. Recently, multi-objective based sparse unmixing methods have presented promising performance because of their advantages in addressing combinatorial problems. A spectral and multi-objective based sparse unmixing (SMoSU) algorithm was proposed in our previous work, which solves the decision-making problem well. However, it does not show outstanding advantages in strong noise cases. To solve the problem, in this paper, SMoSU is improved based on the estimation of distribution algorithms (EDAs). The machine learning based EDAs have been a reliable approach in solving multi-objective problems. However, most of them are for special problems and relatively weak in theoretical foundations. Thus, it is unreliable to extend it directly to sparse unmixing. Here, we improve EDA on the basis of classification and propose a classification-based model for individual generating under the framework of SMoSU (CM-MoSU). In CM-MoSU, the whole population is divided to be positive and negative. Then, the macroinformation of positive individuals is used to guide the generation of new individuals. Therefore, the optimization task could pay more attention to the feasible space with high quality. Moreover, some theoretical analyses are presented to prove the reliability of CM-MoSU. In experiments, several state-of-the-art sparse unmixing algorithms are compared. Both synthetic and real-world experiments demonstrate the effectiveness of CM-MoSU.

**Index Terms**—Classification model, hyperspectral images, multi-objective, sparse unmixing.

## I. INTRODUCTION

**H**YPERSPECTRAL imagery (HSI) could provide a continuous spectrum for a single pixel, which contains reflectivity values in hundreds of bands [1]. HSI has been

Manuscript received May 11, 2018; revised October 25, 2018, March 3, 2019 and June 6, 2019; accepted July 7, 2019. Date of publication August 20, 2019; date of current version November 25, 2019. This work was supported in part by the National Key R&D Program of China under Grant 2017YFC1405605, in part by the National Natural Science Foundation of China under Grant 61671037, in part by the Beijing Natural Science Foundation under Grant 4192034, and in part by the National Defense Science and Technology Innovation Special Zone Project. (*Corresponding author: Zhenwei Shi.*)

X. Xu, Z. Shi, and B. Pan are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, also with the Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with the State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: xuxia@buaa.edu.cn; shizhenwei@buaa.edu.cn; panbin@buaa.edu.cn).

X. Li is with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: xuelong\_li@nwpu.edu.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2928021

an important component of remote sensing in recent years. The advantage of high spectral resolution makes it widely used in many fields, such as land cover classification [2], [3] and target detection [4]. However, due to the complex hyperspectral imaging process and limited spatial resolution, it is likely that each pixel in HSI is a mixture of several different materials [5]. Therefore, exploring and analyzing these compositions are meaningful tasks for many HSI applications.

Hyperspectral unmixing aims at modeling the spectrum of each pixel as a combination of several pure spectra (*endmembers*) and computing the fractions of these endmembers (*abundances*) [6]. Although a variety of models have been proposed successively, linear unmixing model (LMM) is still representative and widely used [7]–[9]. It is easily implemented and flexible, since the interaction among different substances and some other disturbances in the imaging process are not taken into account. During the past years, a tremendous amount of effort has been made to solve the spectral unmixing problem under LMM. Early research is mainly based on geometry [10]–[12], statistics [13], [14], and nonnegative matrix factorization [15]–[18]. However, such methods usually need to assume the existence of at least one pure pixel per endmember, which is difficult to be satisfied in practice [19].

Sparse-regression-based unmixing attracts great interest to researchers since proposed [20]. It is a semisupervised approach that depends neither on the presence of pure pixels nor the capacity of endmember extraction algorithms. Sparse unmixing is to reconstruct the original image using a few pure spectra in a spectral library. The library is composed of many pure spectra collected by field spectroradiometer [21]. Mathematically, sparse unmixing is a  $\ell_0$  norm-based combinatorial optimization problem that is NP-hard. For a hyperspectral image  $\mathbf{Y} \in \mathbb{R}^{L \times N}$  with  $N$  pixels, the LMM is

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{L \times M}$  is the spectral library with  $M$  pure spectra,  $\mathbf{X} \in \mathbb{R}^{M \times N}$  is the abundance matrix of all the  $N$  pixels,  $\mathbf{N} \in \mathbb{R}^{L \times N}$  is the corresponding error term. Due to the physical meaning of sparse unmixing, the following abundance nonnegative constraint (ANC) and abundance sum-to-one constraint (ASC) need to be considered in the model:

$$\begin{aligned} \text{ANC: } \mathbf{X} &= [\mathbf{x}_1, \dots, \mathbf{x}_N] \geq 0 \\ \text{ASC: } \forall \mathbf{x}_i \in \mathbf{X}, \mathbf{1}^T \mathbf{x}_i &= 1, i = 1, \dots, N \end{aligned} \quad (2)$$

where ANC constrains a nonnegative proportion for each endmember, ASC expects a complete decomposition for

each pixel. Then, the optimization problem of sparse unmixing can be written as

$$\begin{aligned} \min_{\mathbf{X} \geq 0} & \|\mathbf{X}\|_{\text{row}-0} \\ \text{s.t.} & \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\text{F}} \leq \delta \end{aligned} \quad (3)$$

where  $\|\mathbf{X}\|_{\text{row}-0}$  (row- $\ell_0$  quasi-norm [22]) denotes the number of nonzero rows in  $\mathbf{X}$ . The ASC is not added in (3) since that ANC could give an automatic imposition to generalized ASC [20].

Recently, many methods have been proposed to solve the sparse unmixing problem. One of the mainly used approaches is approximating the  $\ell_0$  norm. Among them,  $\ell_1$  norm is the most commonly used approximation approach. Quite a few studies have been proposed based on the  $\ell_1$  norm relaxation [23], [24]. In [25]–[27], total variation regularization, nonlocal spatial information, and collaborative sparsity among neighboring pixels were used to exploit similar patterns and structures presented in HSI [28]. In [29], the mismatch between the spectral library and actual spectral signatures was considered to solve the unmixing effect of time-varying and diverse physical conditions in hyperspectral imaging process [30].

Another approach is to solve  $\ell_0$  norm-based sparse unmixing problem directly. In [22] and [31], orthogonal matching pursuit and its improved versions were proposed. Endmembers are selected iteratively from the spectral library in these cases. Multi-objective optimization has become an efficient way to solve the sparse unmixing problem (abbreviated as MoSU). It has also been widely used in band selection [32], [33] and clustering [34]. For sparse unmixing, a triobjective sparse unmixing [35] was proposed and solved based on a cooperative coevolutionary strategy. There are several other MoSU algorithms that encode the spectral library as a binary vector, whose size is the same as the spectra number in the library. Then, sparse unmixing amounts to find the optimal binary vector. In [36], a two-phase MoSU algorithm was proposed. The two biobjective optimization problems are used to determine the endmembers and compute the abundances, respectively. In [37], the typical nondominated sorting genetic algorithm-II is improved to solve the MoSU problem. A drawback of MoSU is that the nondominated sorting strategy is time-consuming. Besides, it may bring extra error when selecting the final solution from the nondominated solution set (the decision-making problem of MoSU). Spectra information-based MoSU is proposed in our previous work [38] to solve these problems. In spectral and multi-objective based sparse unmixing (SMoSU), the multi-objective evolutionary algorithm based on decomposition (MOEA/D) framework is used to obtain a faster evolutionary speed. The addition of spectral information is to solve the decision-making problem for MoSU.

SMoSU is able to find the correct endmembers when the image noise is weak. However, its performance will be significantly impaired by strong noise. This is because SMoSU is heavily dependent on the spectral correlations between selected endmembers and the image data, while noise would weaken these corrections. An available approach to address

this problem is creating a model to represent the distribution of individuals.

In this paper, SMoSU is improved based on the estimation of distribution algorithms (EDAs). EDAs are different from traditional genetic operator-based evolutionary algorithms. They learn a probabilistic model associated with the most outstanding individuals found so far [39]. Then, new solutions are produced by sampling from the probabilistic model. Thus, EDA could use the evolutionary trend of superior solutions and model the evolution process macroscopically [40], [41]. Recently, EDAs have achieved good results in many fields. However, most of them are designed especially for different applications and lack theoretical foundation [42]. Thus, it is unreliable to extend it directly to sparse unmixing.

In the proposed method, a classification-based model (CM) associated with the most promising solutions found so far is developed for individual generating under the framework of SMoSU (termed as CM-MoSU for short). This work is inspired by Yu *et al.* [42], where a classification model was proposed to solve derivative-free problems. Taking the advantages that EDAs could grasp the evolution trend of better solutions by modeling them statistically, a classification-based model is proposed in CM-MoSU. In CM-MoSU, the whole population is divided into two categories: individuals that have better performance are denoted to be positive and negative otherwise. Then, CM is designed as a function of them to generate new individuals. The positive solutions are used to provide active guidance of the search task, and the negative ones are used to maintain the diversity of individuals. To further enhance the individual diversity, the bitwise flipping strategy is adopted. Different from our precious papers [37], [38], in CM-MoSU, it not only plays a role in the evolutionary process but also works as a component in CM. Moreover, some theoretical analyses are provided to prove the reliability of CM-MoSU. Mainly, the failing probability of generating individuals from CM is bounded. The overall failing probability of CM-MoSU is also bounded and proved to be decreased as the iteration progresses. The major contributions of CM-MoSU can be summarized as follows.

- 1) A novel classification-based model for individual generating is proposed in CM-MoSU. This strategy could take advantage of the most promising solutions found so far. Thus, high quality of newly generated individuals could be ensured based on the positive guidance of these solutions.
- 2) Detailed theoretical analysis is presented to support the reliability of CM-MoSU. The failing probability of generating new individuals by the proposed classification model is described quantitatively.

The rest of this paper is organized as follows. Section II introduces the background of multi-objective optimization and SMoSU. In Section III, detailed description of CM-MoSU is presented. Experimental results and analysis are shown in Section IV. Finally, the conclusion is drawn in Section V.

## II. BACKGROUND

In this section, we first review the multi-objective optimization, then the related work SMoSU is briefly introduced.

### A. Multi-Objective Optimization

Before the review of multi-objective optimization, we first introduce evolutionary algorithms. They are inspired by the evolution of species in nature. The species change constantly according to the evolution of individuals. Evolutionary algorithms use this concept to solve a series of technical problems. They define a solution of the problem as an individual. Several individuals form a population, which is the variable of evolutionary algorithms. The population is updated gradually according to certain operators. This population-based strategy makes several solutions be optimized simultaneously in a single iteration. Thus it is a perfect choice to solve the multi-objective optimization problems. Some mathematical definition of multi-objective optimization is introduced as follows.

*Definition 1 (Multi-Objective Optimization Problem):* Let  $\mathbf{u} = (u_1, \dots, u_m)$  be the decision variables. Multi-objective optimization problem aims at finding a solution set such that

$$\begin{aligned} \min F(\mathbf{u}) &= [f_1(\mathbf{u}), f_2(\mathbf{u}), \dots, f_n(\mathbf{u})] \\ \text{s.t. } \mathbf{u} &\in \Omega \subseteq \mathbb{R}^m \end{aligned} \quad (4)$$

where  $\Omega$  is the decision space,  $F : \Omega \rightarrow \mathbb{R}^n$  consists of the  $n$  objective functions.

*Definition 2 (Pareto Dominance):* Suppose  $\mathbf{u}^1$  and  $\mathbf{u}^2$  are two feasible solutions of (4),  $\mathbf{u}^1$  dominates  $\mathbf{u}^2$  ( $\mathbf{u}^1 \succ \mathbf{u}^2$ ) if and only if

$$\begin{aligned} \forall i \in \{1, 2, \dots, n\} : f_i(\mathbf{u}^1) &\leq f_i(\mathbf{u}^2) \wedge \\ \exists j \in \{1, 2, \dots, n\} : f_j(\mathbf{u}^1) &< f_j(\mathbf{u}^2) \end{aligned} \quad (5)$$

*Definition 3 (Pareto Optimal):* A solution  $\mathbf{u}^* \in \Omega$  is Pareto optimal to (4) if and only if  $\nexists \mathbf{u} \in \Omega, \mathbf{u} \succ \mathbf{u}^*$ .

*Definition 4 (Pareto Set (PS)):* PS contains all the Pareto optimal solutions:  $\{\mathbf{u}^* \mid \nexists \mathbf{u} \in \Omega, \mathbf{u} \succ \mathbf{u}^*\}$ .

*Definition 5 (Pareto Front):* Pareto front contains all the Pareto optimal objective vectors:  $\{F(\mathbf{u}^*) \mid \mathbf{u}^* \in \text{PS}\}$ .

In multi-objective optimization, individual evolution usually contains two steps: 1) generating new individuals and 2) selecting better individuals to update the population. Commonly used individual selection methods are classified into two categories: Pareto criterion and non-Pareto criterion. The former is based on the definition of Pareto dominance. However, it is time-consuming to sort individuals. MOEA/D is one of the non-Pareto criterion-based methods. It decomposes a multi-objective problem to several scalar single-objective optimization subproblems, and the objective of each subproblem is a weighted aggregation of the individual objectives.

### B. SMoSU

SMoSU was proposed under the framework of MOEA/D. In SMoSU, spectral information was added to the Tchebycheff decomposition function to solve the decision-making problem. Each spectrum in the spectral library is represented by a binary code. ‘‘1’’ indicates that the spectrum in the corresponding location of the library is selected as an endmember, and ‘‘0’’ otherwise. Thus, the whole spectral library is encoded as a binary vector, whose dimension is equal to the number of

spectra in the library. Accordingly, finding the optimal subset of the library amounts to finding the corresponding optimal binary vector. The optimization problem of endmember selection in SMoSU is

$$\begin{aligned} \min_{\mathbf{s} \in \{0,1\}^m} F(\mathbf{s}) &= [f_1(\mathbf{s}), f_2(\mathbf{s})]^T \\ f_1(\mathbf{s}) &= \begin{cases} +\infty, & \|\mathbf{s}\|_1 = 0 \text{ or } \geq 2k \\ \|\mathbf{Y} - \mathbf{A}_s \mathbf{X}_s\|_F, & \text{otherwise} \end{cases} \\ f_2(\mathbf{s}) &= \left| \|\mathbf{s}\|_1 - k \right| \end{aligned} \quad (6)$$

where  $\mathbf{s} \in \mathbb{R}^M$  is the binary representation of  $\mathbf{A}$ ,  $\|\mathbf{s}\|_1$  denotes the number of endmembers represented by a certain individual,  $\mathbf{A}_s$  is the corresponding subset of  $\mathbf{A}$ ,  $\mathbf{X}_s$  is the abundances of endmembers,  $k$  is the actual endmember number in hyperspectral image  $\mathbf{Y}$ ,  $f_1(\mathbf{s})$  is the reconstruction error, and  $f_2(\mathbf{s})$  is the sparsity error.  $f_1(\mathbf{s})$  is meaningful when  $0 < \|\mathbf{s}\|_1 < 2k$  to exclude trivial or overly bad solutions. It is worth noticing that  $k$  is usually unknown in practical application; therefore, HySime method is used to give it an estimation [43]. After decomposition, the  $i$ th subproblem in SMoSU is

$$\begin{aligned} \min_{\mathbf{s}_i} g_i^{\text{st}}(\mathbf{s}_i \mid \lambda_i, \mathbf{z}^*, \mathbf{s}^*), \quad \text{s.t. } 0 < \|\mathbf{s}_i\|_1 &< 2k \\ \text{where} \\ g_i^{\text{st}}(\mathbf{s}_i \mid \lambda_i, \mathbf{z}^*, \mathbf{s}^*) &= \max_{1 \leq j \leq 2} \{ \lambda_i^j |f_j(\mathbf{s}_i) - z_j^*| \} + \mu \cdot \text{SID}(\mathbf{A}_{\mathbf{s}_i}, \mathbf{A}_{\mathbf{s}^*}) \end{aligned} \quad (7)$$

$\mathbf{z}^*$  is the ideal point that is closest to the origin,  $g_i^{\text{st}}(\mathbf{s}_i \mid \lambda_i, \mathbf{z}^*, \mathbf{s}^*)$  is the weighted Tchebycheff distance of individual  $\mathbf{s}_i$  to the ideal point  $\mathbf{z}^*$ ,  $\mathbf{s}^*$  is the binary vector of  $\mathbf{z}^*$ ,  $\mathbf{A}_{\mathbf{s}_i}$  and  $\mathbf{A}_{\mathbf{s}^*}$  contain the corresponding spectra of individual  $\mathbf{s}_i$  and  $\mathbf{s}^*$ ,  $\text{SID}(\mathbf{A}_{\mathbf{s}_i}, \mathbf{A}_{\mathbf{s}^*})$  is the spectral information divergence of  $\mathbf{A}_{\mathbf{s}_i}$  and  $\mathbf{A}_{\mathbf{s}^*}$ ,  $\mu$  is a regularization parameter.

In SMoSU, a bitwise flipping strategy is used to generate new individuals. This strategy can maintain the diversity of individuals to some extent. However, it is somewhat reckless and isolated, since no contextual information of the evolutionary process is used. Moreover, as the spectral correlation information is heavily dependent on the imaging quality, SMoSU cannot achieve satisfying performance when the image noise is strong.

## III. CM-MoSU

To solve the problems in SMoSU, a classification-based model for individual generation is designed to improve SMoSU in this paper. In this section, the proposed model CM is described first. Then, the framework of CM-MoSU is introduced. At last, theoretical analysis of CM-MoSU is presented.

### A. CM

In order to draw on the evolutionary information of superior individuals, CM-MoSU divides the population into two categories: individuals with better performance are denoted to be positive, and the remaining inferior individuals are denoted to be negative. Then, new individuals are generated according to

a linear model of them. This classification-based model (CM) is defined as

$$\mathbf{s}_{\text{cm}} = \mathbf{s}_{\text{pos}} \odot \mathbf{d}_{\text{pos}} + \mathbf{s}_{\text{neg}} \odot \mathbf{d}_{\text{neg}} + \mathbf{s}_{\text{bit}} \odot \mathbf{d}_{\text{bit}} \quad (8)$$

where  $\mathbf{s}_{\text{cm}}$  is the new individual generated by CM,  $\mathbf{s}_{\text{pos}}$  and  $\mathbf{s}_{\text{neg}}$  are two randomly selected individuals from the positive class and negative class, respectively,  $\mathbf{s}_{\text{bit}}$  is the bitwise flipping product of  $\mathbf{s}$ .  $\mathbf{d}_{\text{pos}}$ ,  $\mathbf{d}_{\text{neg}}$  and  $\mathbf{d}_{\text{bit}}$  are their corresponding coefficients, which have the same dimension with them.  $\mathbf{s}_{\text{pos}} \odot \mathbf{d}_{\text{pos}}$  is the elementwise product of  $\mathbf{s}_{\text{pos}}$  and  $\mathbf{d}_{\text{pos}}$ , and so on.

It can be observed that the core of CM is the determination of  $\mathbf{d}_{\text{pos}}$ ,  $\mathbf{d}_{\text{neg}}$  and  $\mathbf{d}_{\text{bit}}$  which are learned from the positive and negative individuals. In the following, a detailed description of these three parameters is presented, respectively.

- 1)  $\mathbf{d}_{\text{pos}}$  is used to take out the outstanding information of  $\mathbf{s}_{\text{pos}}$ . In  $\mathbf{s}_{\text{pos}}$ , elements that are different from all the negative individuals are considered as preponderant. The corresponding locations in  $\mathbf{d}_{\text{pos}}$  are set as “1” (“0” for the remaining locations). Thus, information of these elements could be retained when generating new individuals.
- 2)  $\mathbf{d}_{\text{neg}}$  is used to maintain the diversity of individuals. Diverse solutions are essential in multi-objective optimization, which is used to avoid getting stuck in a local optimum. In CM-MoSU, one of the nonzero elements in  $\mathbf{d}_{\text{pos}}$  is selected randomly and set to be “0.” Accordingly, the same location in  $\mathbf{d}_{\text{neg}}$  is set as “1” (“0” for the others). Information of only one element in  $\mathbf{s}_{\text{neg}}$  is retained. The reason is that if we set many elements of  $\mathbf{d}_{\text{neg}}$  to “1,” the superiority of  $\mathbf{s}_{\text{pos}}$  may be overshadowed.
- 3)  $\mathbf{d}_{\text{bit}}$  is used to improve the diversity of individuals, so that a more extensive search in the feasible region could be obtained. The locations at which  $\mathbf{d}_{\text{pos}}$  and  $\mathbf{d}_{\text{neg}}$  are both “0,” are set as “1” for  $\mathbf{d}_{\text{bit}}$ . In fact, the information of positive and negative individuals is not used for them. Thus, it is natural to add some other information to improve the performance of the new solution. Taking the advantage that bitwise flipping could produce new individuals diversely, the values of  $\mathbf{s}_{\text{bit}}$  at these locations are used.

To make a clear introduction for CM, an example is presented. Assume positive individuals as  $\{[00101], [11001]\}$  and negative individuals as  $\{[10001], [01011]\}$ . Select an individual  $\mathbf{s}_{\text{pos}} = [00101]$  from the positive set. In  $\mathbf{s}_{\text{pos}}$ , the first four elements are different from the negative set. They are preponderant for  $\mathbf{s}_{\text{pos}}$ . Thus, the coefficient  $\mathbf{d}_{\text{pos}}$  is set as  $[11110]$ . Then, one of the nonzero elements of  $\mathbf{d}_{\text{pos}}$  is flipped, and  $\mathbf{d}_{\text{pos}}$  could be  $[10110]$ . Accordingly,  $\mathbf{d}_{\text{neg}} = [01000]$  and  $\mathbf{d}_{\text{bit}} = [00001]$ .

## B. Framework of CM-MoSU

The flowchart of CM-MoSU is shown in Fig. 1.

*Step 1:* Determine the dimension of the binary variable  $\mathbf{s}$  according to the spectra number of the library  $\mathbf{A} \in \mathbb{R}^{L \times M}$ . Initialize a population  $\mathbf{P} = [\mathbf{s}^1, \dots, \mathbf{s}^p]$ , where  $\mathbf{s}^i \in \mathbb{R}^M$  ( $i = 1, \dots, p$ ) is the  $i$ th individual and  $p$  is the population size. Initialize the number of positive individuals (positive

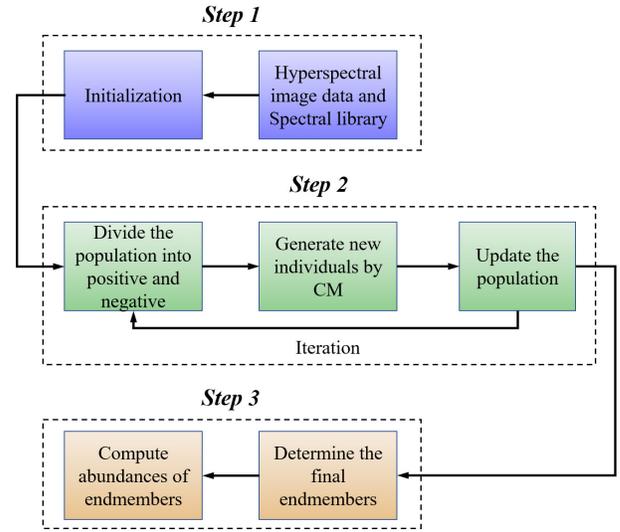


Fig. 1. Flowchart of CM-MoSU.

size) as  $p_{\text{pos}}$ . Variables of the SMOsU framework are set as the default.

*Step 2:* For each individual  $\mathbf{s}^i$ ,  $\mathbf{s}_{\text{bit}}^i$  is generated based on the bitwise flipping strategy. Sort all individuals roughly according to the Euclidean distance from each individual’s objective values to the origin. The first  $p_{\text{pos}}$  individuals that are closer to (0, 0) are denoted to be positive and negative otherwise. Select an individual  $\mathbf{s}_{\text{pos}}^i$  from the positive individuals, and an individual  $\mathbf{s}_{\text{neg}}^i$  from the negative individuals. Determine the coefficients  $\mathbf{d}_{\text{pos}}$ ,  $\mathbf{d}_{\text{neg}}$ , and  $\mathbf{d}_{\text{bit}}$  for  $\mathbf{s}_{\text{pos}}$ ,  $\mathbf{s}_{\text{neg}}$ , and  $\mathbf{s}_{\text{bit}}$  according to Section III-A, and generate a new individual  $\mathbf{s}_{\text{cm}}$  based on CM. At last, the offspring of  $\mathbf{s}^i$  is set as  $\mathbf{s}_{\text{cm}}$  with a probability  $\lambda$  and  $\mathbf{s}_{\text{bit}}$  with a probability  $1 - \lambda$ . The setting of  $\lambda$  is analyzed in Section IV.

*Step 3:* Update each individual by using the improved Tchebycheff decomposition approach in SMOsU, where a single solution was obtained finally. Then, we do not need to choose the final solution manually from the nondominated solution set.

*Step 4:* Stop if the stopping criteria is satisfied. Return the solution in  $\mathbf{P}$  as  $\mathbf{s}^*$ . Take out the nonzero indexes in  $\mathbf{s}^*$  and find out the corresponding spectra  $\mathbf{A}_{\mathbf{s}^*}$  from  $\mathbf{A}$ . The spectra in  $\mathbf{A}_{\mathbf{s}^*}$  are endmembers of the hyperspectral image  $\mathbf{Y}$ . Compute abundances  $\mathbf{X}_{\mathbf{s}^*}$  of the endmembers based on nonnegative least squares algorithm:  $\mathbf{X}_{\mathbf{s}^*} = \arg \min_{\mathbf{X}_{\mathbf{s}^*} \geq 0} \|\mathbf{Y} - \mathbf{A}_{\mathbf{s}^*} \mathbf{X}_{\mathbf{s}^*}\|_F$ .

## C. Theoretical Analysis of CMoSU

In the following section, the theoretical analysis of CMoSU is presented.

*Lemma 1:* Let  $S$  be a solution space, and  $D$  be a subset of  $S$ . Denote  $\mathcal{U}_S$  as the uniform distribution over  $S$ . Then, generating new individuals from the bitwise flipping strategy can be seen as sampling from  $\mathcal{U}_S$ . Define  $\#D = \sum_{\mathbf{s} \in S} \mathbb{I}[x \in D]$  and  $|D| = \#D/\#S$ , where  $\mathbb{I}[\cdot]$  is the indicator function. Let  $D_\epsilon = \{\mathbf{s} \in S | f_1(\mathbf{s}) - f_1(\mathbf{s}^*) \leq \epsilon\}$  for  $\epsilon > 0$ ,  $D_\delta = \{\mathbf{s} \in S | f_2(\mathbf{s}) - f_2(\mathbf{s}^*) \leq \delta\}$  for  $\delta > 0$ , and  $D_{\epsilon, \delta} = D_\epsilon \cap D_\delta = \{\mathbf{s} \in S | \mathcal{F}(\mathbf{s}) - \mathcal{F}(\mathbf{s}^*) \leq [\epsilon, \delta]\}$ . Suppose  $h_t$  is a classifier

that dividing  $S$  to be positive and negative at the  $t$ th iteration.  $D_{h_t} = \{\mathbf{s} \in S | h_t(\mathbf{s}) = +1\}$  is used to represent the positive region of  $S$  accordingly, where the first  $p_{\text{pos}}$  individual are located. Then,  $|D_{h_t}|$  is bounded by

$$|D_{\epsilon,\delta}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t} \leq |D_{h_t}| \leq |D_{\epsilon,\delta}| + \mathcal{R}_{\epsilon,t} + \mathcal{R}_{\delta,t} \quad (9)$$

and  $|D_{\epsilon,\delta} \cap D_{h_t}|$  is bounded by

$$|D_{\epsilon,\delta} \cap D_{h_t}| \geq |D_{\epsilon,\delta}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t} \quad (10)$$

where  $\mathcal{R}_{\epsilon,t} = |D_{\epsilon} \Delta D_{h_t}|$  and  $\mathcal{R}_{\delta,t} = |D_{\delta} \Delta D_{h_t}|$  are the generation error of  $h_t$  with respect to  $D_{\epsilon}$  and  $D_{\delta}$  under the distribution  $\mathcal{U}_S$ .

*Proof:* 1) *Bounding  $|D_{h_t}|$ :* Denote  $\Delta$  as a symmetric difference operator, we have

$$\||D_{\epsilon,\delta}| - |D_{h_t}|\| \leq |(D_{\epsilon} \cap D_{\delta}) \Delta D_{h_t}| \leq |D_{\epsilon} \Delta D_{h_t}| + |D_{\delta} \Delta D_{h_t}| \quad (11)$$

According to the definition of  $\mathcal{R}_{\epsilon,t}$  and  $\mathcal{R}_{\delta,t}$ ,  $|D_{h_t}|$  could be bounded by

$$|D_{\epsilon,\delta}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t} \leq |D_{h_t}| \leq |D_{\epsilon,\delta}| + \mathcal{R}_{\epsilon,t} + \mathcal{R}_{\delta,t} \quad (12)$$

2) *Bounding  $|D_{\epsilon,\delta} \cap D_{h_t}|$ :*

$$\begin{aligned} |D_{\epsilon,\delta} \cap D_{h_t}| &= |D_{\epsilon,\delta} \cup D_{h_t}| - |D_{\epsilon,\delta} \Delta D_{h_t}| \\ &\geq |D_{\epsilon,\delta} \cup D_{h_t}| - |D_{\epsilon} \Delta D_{h_t}| - |D_{\delta} \Delta D_{h_t}| \\ &= |D_{\epsilon,\delta} \cup D_{h_t}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t} \\ &\geq |D_{\epsilon,\delta}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t} \end{aligned} \quad (13)$$

**Theorem 1:** Let  $\mathcal{U}_{h_t}$  be the uniform distribution over  $D_{h_t}$ , and  $\mathcal{T}_{h_t}$  be the probability distribution defined on  $D_{h_t}$ . Then generating new individuals from CM can be approximated as sampling from  $\mathcal{T}_{h_t}$ , since most information of  $\mathbf{s}_{\text{cm}}$  comes from positive individuals. Denote  $D_{KL}$  the Kullback–Leibler divergence (i.e., relative entropy) between two probability distributions.  $P_{\mathcal{T}_{h_t}}$  is the failing probability of individual generating from  $D_{h_t}$ . The generation fails if the solution  $\tilde{\mathbf{s}}$  satisfies  $f_1(\tilde{\mathbf{s}}) - f_1(\mathbf{s}^*) > \epsilon$  or  $f_2(\tilde{\mathbf{s}}) - f_2(\mathbf{s}^*) > \delta$ . Then,  $P_{\mathcal{T}_{h_t}}$  is upper bounded by

$$P_{\mathcal{T}_{h_t}} \leq \frac{2\Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} + 2\Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}}}{|D_{\epsilon,\delta}| + \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} + \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}}} \quad (14)$$

where  $\Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} = \mathcal{R}_{D_t}^{\epsilon} + \#(D_{\epsilon} \Delta D_{h_t}) (\frac{1}{2} D_{KL}(D_t || \mathcal{U}_S))^{1/2}$  and  $\Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}} = \mathcal{R}_{D_t}^{\delta} + \#(D_{\delta} \Delta D_{h_t}) (\frac{1}{2} D_{KL}(D_t || \mathcal{U}_S))^{1/2}$  are the upper bound of  $\mathcal{R}_{\epsilon,t}$  and  $\mathcal{R}_{\delta,t}$ , respectively,  $\mathcal{R}_{D_t}^{\epsilon}$  and  $\mathcal{R}_{D_t}^{\delta}$  are the generation error under distribution  $D_t$  with respect to  $D_{\epsilon}$  and  $D_{\delta}$ , respectively.  $D_t = \lambda \mathcal{T}_{h_t} + (1 - \lambda) \mathcal{U}_S$  is the individual generation distribution at iteration  $t$ .

*Proof:* 1) *Bounding  $\mathcal{R}_{\epsilon,t}$  and  $\mathcal{R}_{\delta,t}$ :* Actually, the individual generation in the classification-model-based operator is based

on distribution  $D_t$  in each iteration, then we have

$$\begin{aligned} \mathcal{R}_{D_t}^{\epsilon} &= \sum_{\mathbf{s} \in S} D_t \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon} \Delta D_{h_t}] \\ &= \sum_{\mathbf{s} \in S} (\mathcal{U}_S + D_t - \mathcal{U}_S) \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon} \Delta D_{h_t}] \\ &= \mathcal{R}_{\epsilon,t} + \sum_{\mathbf{s} \in S} (D_t - \mathcal{U}_S) \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon} \Delta D_{h_t}] \\ &\geq \mathcal{R}_{\epsilon,t} - \sum_{\mathbf{s} \in S} \sup |D_t - \mathcal{U}_S| \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon} \Delta D_{h_t}] \\ &\geq \mathcal{R}_{\epsilon,t} - \#(D_{\epsilon} \Delta D_{h_t}) \sqrt{\frac{1}{2} D_{KL}(D_t || \mathcal{U}_S)} \end{aligned} \quad (15)$$

In a similar way,  $\mathcal{R}_{D_t}^{\delta}$  could be also bounded by

$$\mathcal{R}_{D_t}^{\delta} \geq \mathcal{R}_{\delta,t} - \#(D_{\delta} \Delta D_{h_t}) \sqrt{\frac{1}{2} D_{KL}(D_t || \mathcal{U}_S)} \quad (16)$$

Thus,

$$\begin{aligned} \mathcal{R}_{\epsilon,t} &\leq \mathcal{R}_{D_t}^{\epsilon} + \#(D_{\epsilon} \Delta D_{h_t}) \sqrt{\frac{1}{2} D_{KL}(D_t || \mathcal{U}_S)} = \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} \\ \mathcal{R}_{\delta,t} &\leq \mathcal{R}_{D_t}^{\delta} + \#(D_{\delta} \Delta D_{h_t}) \sqrt{\frac{1}{2} D_{KL}(D_t || \mathcal{U}_S)} = \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}} \end{aligned} \quad (17)$$

2) *Bounding  $P_{\mathcal{T}_{h_t}}$ :*

$$\begin{aligned} P_{\mathcal{T}_{h_t}} &= 1 - \sum_{\mathbf{s} \in D_{h_t}} \mathcal{T}_{h_t} \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon,\delta}] \\ &= 1 - \sum_{\mathbf{s} \in D_{h_t}} (\mathcal{T}_{h_t} - \mathcal{U}_{h_t} + \mathcal{U}_{h_t}) \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon,\delta}] \\ &= 1 - \frac{|D_{\epsilon,\delta} \cap D_{h_t}|}{|D_{h_t}|} - \sum_{\mathbf{s} \in D_{h_t}} (\mathcal{T}_{h_t} - \mathcal{U}_{h_t}) \cdot \mathbb{I}[\mathbf{s} \in D_{\epsilon,\delta}] \\ &\leq 1 - \frac{|D_{\epsilon,\delta} \cap D_{h_t}|}{|D_{h_t}|} + \sqrt{\frac{1}{2} D_{KL}(\mathcal{T}_{h_t} || \mathcal{U}_{h_t})} \cdot \sum_{\mathbf{s} \in D_{h_t}} \mathbb{I}[\mathbf{s} \in D_{\epsilon,\delta}] \\ &= 1 - \frac{|D_{\epsilon,\delta} \cap D_{h_t}|}{|D_{h_t}|} + \#(D_{\epsilon,\delta} \cap D_{h_t}) \sqrt{\frac{1}{2} D_{KL}(\mathcal{T}_{h_t} || \mathcal{U}_{h_t})} \end{aligned} \quad (18)$$

According to the bounds of  $|D_{h_t}|$ ,  $|D_{\epsilon,\delta} \cap D_{h_t}|$ ,  $\mathcal{R}_{\epsilon,t}$  and  $\mathcal{R}_{\delta,t}$ , we have

$$\begin{aligned} P_{\mathcal{T}_{h_t}} &\leq 1 - \frac{|D_{\epsilon,\delta}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t}}{|D_{\epsilon,\delta}| + \mathcal{R}_{\epsilon,t} + \mathcal{R}_{\delta,t}} \\ &\quad + \#(D_{\epsilon,\delta} \cap D_{h_t}) \sqrt{\frac{1}{2} D_{KL}(\mathcal{T}_{h_t} || \mathcal{U}_{h_t})} \end{aligned} \quad (19)$$

Since  $\mathcal{T}_{h_t} = \mathcal{U}_{h_t}$ , we get  $D_{KL}(\mathcal{T}_{h_t} || \mathcal{U}_{h_t}) = 0$ . Thus

$$\begin{aligned} P_{\mathcal{T}_{h_t}} &\leq 1 - \frac{|D_{\epsilon,\delta}| - \mathcal{R}_{\epsilon,t} - \mathcal{R}_{\delta,t}}{|D_{\epsilon,\delta}| + \mathcal{R}_{\epsilon,t} + \mathcal{R}_{\delta,t}} \\ &\leq 1 - \frac{|D_{\epsilon,\delta}| - \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} - \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}}}{|D_{\epsilon,\delta}| + \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} + \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}}} \\ &= \frac{2\Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} + 2\Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}}}{|D_{\epsilon,\delta}| + \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\epsilon}} + \Psi_{D_{KL}}^{\mathcal{R}_{D_t,\delta}}} \end{aligned} \quad (20)$$

*Lemma 2:* For individual generation,  $P_{\mathcal{U}_S}$  and  $P_{\mathcal{T}_{h_t}}$  are the failing probabilities when generating from  $\mathcal{U}_S$  and  $\mathcal{T}_{h_t}$ , respectively.  $P_{fail}$  is the overall failing probability for the balanced individual generating procedure. Based on these definitions,  $P_{fail}$  is lower bounded by

$$P_{fail} \geq ((4\lambda(1-\lambda))^{\frac{T}{2}} P_{\mathcal{U}_S})^p \prod_{t=1}^T (P_{\mathcal{T}_{h_t}})^{\frac{p}{2}}$$

*Proof:* In each iteration, a new individual is generated from  $\mathcal{U}_S$  or  $\mathcal{T}_{h_t}$  with a probability  $\lambda$ . The generation fails if  $f_1(\tilde{\mathbf{s}}) - f_1(\mathbf{s}^*) > \epsilon$  or  $f_2(\tilde{\mathbf{s}}) - f_2(\mathbf{s}^*) > \delta$ . Thus, the overall failing probability can be expressed as

$$\begin{aligned} P_{fail} &= P(f_1(\tilde{\mathbf{s}}) - f_1(\mathbf{s}^*) > \epsilon \vee f_2(\tilde{\mathbf{s}}) - f_2(\mathbf{s}^*) > \delta) \\ &= P_{\mathcal{U}_S}^p \cdot \prod_{t=1}^T \sum_{i=0}^p C_p^i (1-\lambda)^i \lambda^{p-i} P_{\mathcal{U}_S}^i P_{\mathcal{T}_{h_t}}^{p-i} \\ &= P_{\mathcal{U}_S}^p \cdot \prod_{t=1}^T ((1-\lambda)P_{\mathcal{U}_S} + \lambda P_{\mathcal{T}_{h_t}})^p \\ &\geq P_{\mathcal{U}_S}^p \cdot \prod_{t=1}^T (2\sqrt{\lambda(1-\lambda)P_{\mathcal{U}_S}P_{\mathcal{T}_{h_t}}})^p \\ &= (2^T P_{\mathcal{U}_S})^p \cdot \prod_{t=1}^T (\lambda(1-\lambda)P_{\mathcal{U}_S}P_{\mathcal{T}_{h_t}})^{\frac{p}{2}} \\ &= ((4\lambda(1-\lambda))^{\frac{T}{2}} P_{\mathcal{U}_S})^p \prod_{t=1}^T (P_{\mathcal{T}_{h_t}})^{\frac{p}{2}} \end{aligned} \quad (21)$$

*Theorem 2:* The overall failing probability is decreased with the increase in iteration number.

*Proof:* According to *Lemma 2*,  $P_{fail} = P_{\mathcal{U}_S}^p \cdot \prod_{t=1}^T ((1-\lambda)P_{\mathcal{U}_S} + \lambda P_{\mathcal{T}_{h_t}})^p$ , thus we have

$$\frac{P_{fail}(T+1)}{P_{fail}(T)} = ((1-\lambda)P_{\mathcal{U}_S} + \lambda P_{\mathcal{T}_{h_{T+1}}})^p. \quad (22)$$

To guarantee the monotone decreasing over the iteration number  $T$ ,  $((1-\lambda)P_{\mathcal{U}_S} + \lambda P_{\mathcal{T}_{h_{T+1}}})^p$  should be smaller than 1, that is,

$$(1-\lambda)P_{\mathcal{U}_S} + \lambda P_{\mathcal{T}_{h_{T+1}}} \leq 1 \Rightarrow P_{\mathcal{T}_{h_{T+1}}} \leq \frac{1 + (\lambda-1)P_{\mathcal{U}_S}}{\lambda}. \quad (23)$$

Since

$$\begin{aligned} \frac{1 + (\lambda-1)P_{\mathcal{U}_S}}{\lambda} - 1 &= \frac{(1-\lambda)(1-P_{\mathcal{U}_S})}{\lambda} \geq 0 \\ &\Rightarrow \frac{1 + (\lambda-1)P_{\mathcal{U}_S}}{\lambda} \geq 1 \end{aligned} \quad (24)$$

the inequations in (23) are satisfied. Thus, the failing probability is decreased with the increase in iteration number. ■

*Theorem 3:*  $P_{fail}$  could take its lower bound if  $\lambda \rightarrow 1$ . *Proof:* In (21),  $(1-\lambda)P_{\mathcal{U}_S} + \lambda P_{\mathcal{T}_{h_t}} = 2(\lambda(1-\lambda)P_{\mathcal{U}_S}P_{\mathcal{T}_{h_t}})^{1/2}$  is satisfied when  $(1-\lambda)P_{\mathcal{U}_S} = \lambda P_{\mathcal{T}_{h_t}}$ . Then, we have

$$\lambda = \frac{P_{\mathcal{U}_S}}{P_{\mathcal{T}_{h_t}} + P_{\mathcal{U}_S}}. \quad (25)$$

Since  $0 < P_{\mathcal{T}_{h_t}} < 1$ ,  $\lambda$  is bounded in  $((P_{\mathcal{U}_S}/P_{\mathcal{T}_{h_t}} + P_{\mathcal{U}_S}), 1)$ . Regardless of the value of  $P_{\mathcal{U}_S}$  and  $P_{\mathcal{T}_{h_t}}$ ,  $P_{fail}$  could take its lower bound if  $\lambda \rightarrow 1$ . Thus, in this paper,  $\lambda$  is set as 0.99 for all experiments. ■

## IV. EXPERIMENTS

In this section, three synthetic and one real-world experiments are conducted to validate the performance of CM-MoSU. Synthetic experiments are widely used for sparse unmixing to provide quantitative analysis, since it is hard to obtain the true abundance maps in practice. Although real-world images are also considered, only qualitative analysis is available in these cases. Several state-of-the-art methods, SUnSAL [24], SUnSAL-TV [25], SMP [22], RSFoBa-2, RSFoBa-Inf [31], and SMoSU [38] are taken as comparisons for CM-MoSU. Parameters in the six comparison methods are adapted to find the best results.

In synthetic data 1, we discuss the influence of the important parameter, the number of positive individuals. Furthermore, the computational cost of CM-MoSU is also analyzed. To avoid repetition, these experiments on synthetic experiments 2 and 3 are not shown, because very similar results are observed.

Here, we design a new criterion to evaluate the performance of the endmember selection process.

*The Criterion for Endmember Selection:* Inspired by receiver operating characteristic (ROC), the true positive rate (TPR) and false positive rate (FPR) are introduced to assess the endmember selection results. We also call this criterion as ROC. Since SUnSAL and SUnSAL-TV are not subset selection-based, they are not taken into comparisons here. The TPR and FPR are computed as follows.

- 1) *True Negative (TN):* negative case and predicted negative
- 2) *True Positive (TP):* positive case and predicted positive
- 3) *False Negative (FN):* positive case but predicted negative
- 4) *False Positive (FP):* negative case but predicted positive

$$TPR = \frac{TP}{TP + FN} \quad (26)$$

$$FPR = \frac{FP}{TN + FP}. \quad (27)$$

For the endmember selection problem, active spectra in the spectral library are denoted as positive cases and negative otherwise. A larger TPR indicates that the results of endmember selection contain more active ones. A smaller FPR indicates that less inactive spectra are included in the selection results. In a word, larger TPR and smaller FPR correspond to better results, so the scatters on the upper left corner of ROC figures correspond to better results.

In addition, a popular criterion for abundance estimation is adopted.

*The Criterion for Abundance Estimation:* The signal-to-reconstruction error (SRE) is computed to evaluate the abundance estimations

$$SRE \equiv 10 \log_{10} (E[\|\mathbf{X}\|_F^2] / E[\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2]) \quad (28)$$

where  $\hat{\mathbf{X}}$  is the estimated value. Generally, the larger the SRE is, the more accurate the estimation is.

TABLE I

TPR AND FPR RESULTS OF SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, AND CM-MoSU ON SYNTHETIC DATA 2 WITH 20-dB CORRELATED NOISE

	$k$	3	4	5	6	7	8	9	10
SMP	TPR	1	7.50e-1	6.00e-1	1	8.57e-1	8.89e-1	9.00e-1	1
	FPR	2.00e-3	4.00e-3	0	4.67e-2	3.67e-2	4.29e-2	6.95e-2	2.25e-2
RSFoBa-2	TPR	1	1	1	1	1	1	8.89e-1	1
	FPR	2.00e-3	1.01e-2	2.00e-3	5.28e-2	8.98e-2	9.82e-2	9.82e-2	1.31e-1
RSFoBa-Inf	TPR	1	1	1	1	1	1	1	1
	FPR	2.00e-3	8.10e-3	4.10e-3	2.03e-3	6.52e-2	8.57e-2	6.95e-2	8.40e-2
SMoSU	TPR	1	1	1	1	1	1	1	1
	FPR	4.00e-3	2.00e-3	4.10e-3	6.10e-3	4.10e-3	6.10e-3	6.10e-3	6.10e-3
CM-MoSU	TPR	1	1	1	1	1	1	1	1
	FPR	4.00e-3	2.00e-3	4.10e-3	6.10e-3	4.10e-3	6.10e-3	6.10e-3	6.10e-3

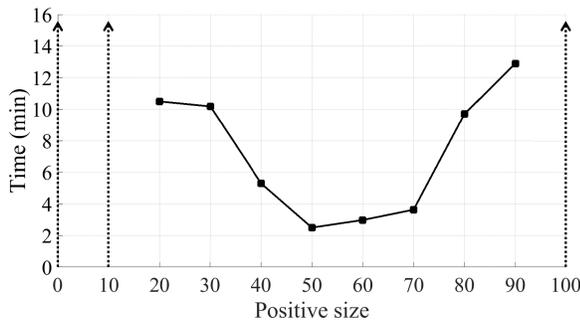


Fig. 2. Processing time of CM-MoSU on synthetic data 1 with  $p_{\text{pos}}$  varying from 0 to 100. The dotted lines denote that under these settings CM-MoSU cannot find the true endmembers exactly.

### A. Synthetic Data 1

The ground-truth spectra of synthetic data 1 are selected from the Chapter 1 of the United States Geological Survey (USGS) digital spectral library (splib06a)<sup>1</sup>, which contains 498 spectral signatures that collected under 224 spectral bands. Synthetic data 1 contains  $64 \times 64$  pixels. The true endmember number  $\hat{k}$  of synthetic data 1 is varying from 3 to 10. Among them, 5 similar spectra (Actinolite HS116.3B, Actinolite HS22.3B, Actinolite HS315.4B, Actinolite NMNH80714, and Actinolite NMNHR16485) are selected specially to increase the unmixing difficulty. To better verify the performance of CM-MoSU, all the images are corrupted by correlated noise, whose signal-to-noise ratios ( $\text{SNR} \equiv 10 \log_{10}(\frac{\|\mathbf{Y}\|_F^2}{\|\mathbf{N}\|_F^2})$ ) are 20/30/40 dB, respectively. Abundances are generated based on Dirichlet distribution and forced to be smaller than 0.7 to make the unmixing process more challenging [38].

1) *Discussion for  $p_{\text{pos}}$  and Computational Time:* To test the influence of  $p_{\text{pos}}$  in CM-MoSU, the convergence time for different  $p_{\text{pos}}$  are shown in Fig. 2. The experiments are conducted with  $p_{\text{pos}}$  varying from 10 to 100 ( $p = 100$  and

the interval is 10). The endmember number  $\hat{k}$  is 5 and the correlated noise level is 20 dB here.

In the conditions  $p_{\text{pos}} = 0/10/100$ , CM-MoSU cannot exactly find the true endmembers, thus the convergence time is not shown. These results demonstrate the effectiveness of the positive/negative individuals:  $p_{\text{pos}} = 0/100$  indicates that no positive/negative individuals are used. The best result is observed when  $p_{\text{pos}}$  is set as 50, i.e., the average distribution of positive and negative individuals could improve the convergence speed. Based on this observation, we set  $p_{\text{pos}}$  as half volume of the original population. It is worth noting that the unbalanced positive/negative individuals' numbers will lead to long time for convergence ( $p_{\text{pos}} = 90$ ) or even nonconvergence ( $p_{\text{pos}} = 10$ ). This result is another evidence for our conclusion of equally setting positive/negative individuals.

Fig. 2 could also be considered as a speed discussion about CM-MoSU. With proper  $p_{\text{pos}}$ , CM-MoSU can generate the unmixing results by about 2 min. Although it looks that CM-MoSU costs a little more time than some single-objective based methods [22], [31], CM-MoSU has its unique advantage. Single-objective based methods (such as [22], [31]) usually have to adjust the regularization parameters for different data sets. This process requires empirical and manual participation. By comparison, CM-MoSU is a completely automatic algorithm which can be directly used for different data sets. In other words, the reported time by CM-MoSU is the total processing time.

2) *Endmember Assessment:* Tables I–III present the TPR and FPR results under 20-/30-/40-dB noise cases, respectively, for SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, and CM-MoSU. From the tables, the TPR and FPR results of SMoSU and CM-MoSU are the same and better than the other methods for each  $\hat{k} = 3-10$  cases. The TPR results of SMP are smaller than 1 in most cases, i.e., SMP loses several active endmembers in these cases. The TPRs of RSFoBa-Inf, SMoSU, and CM-MoSU are all 1 for different  $k$  and different noise level. They could find all active endmembers. The FPR results of SMP, RSFoBa-2, and RSFoBa-Inf are larger than 0 and

<sup>1</sup>Available online: <http://speclab.cr.usgs.gov/spectral-lib.html>

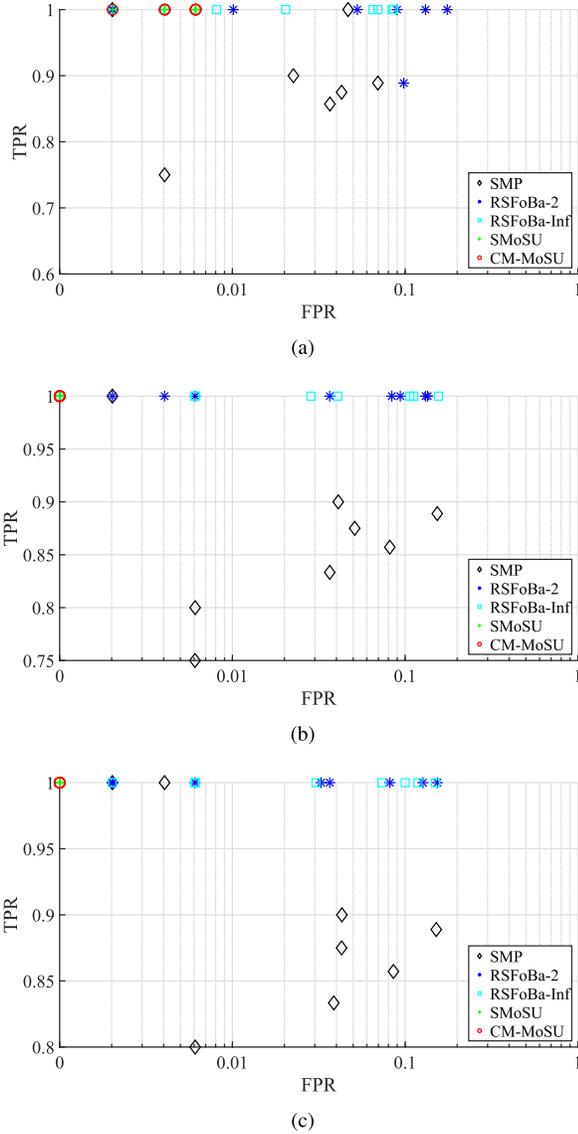


Fig. 3. ROC scatter diagram (logarithmic abscissa) of endmember selection results for SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, and CM-MoSU on synthetic data 1. (a)–(c) Results for 20/30/40 dB data, respectively. Note that although scatters of SMoSU and CM-MoSU are overlapped in some cases, the selected endmembers are not the same (see Fig. 4).

increase with  $k$  increases, i.e., they obtain more redundant endmembers in the larger  $k$  cases. However, the FPRs of SMoSU and CM-MoSU are 0 in all the 30- and 40-dB cases, they find the active endmembers exactly.

Fig. 3 shows the ROC scatter diagram (logarithmic abscissa) of SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, and CM-MoSU under 20-/30-/40-dB noises. Each subfigure contains all the  $\hat{k} = 3-10$  cases. According to (26) and (27), points closer to the upper left represent better results. From Fig. 3, the results of SMP and RSFoBa-2 are farther to the upper left compared with the other algorithms.

Note that although the TPR and FPR results of SMoSU and CM-MoSU are the same, the selected endmembers may be different. This difference would affect the results of abundance inversion, as shown in Fig. 4. TPRs of CM-MoSU,

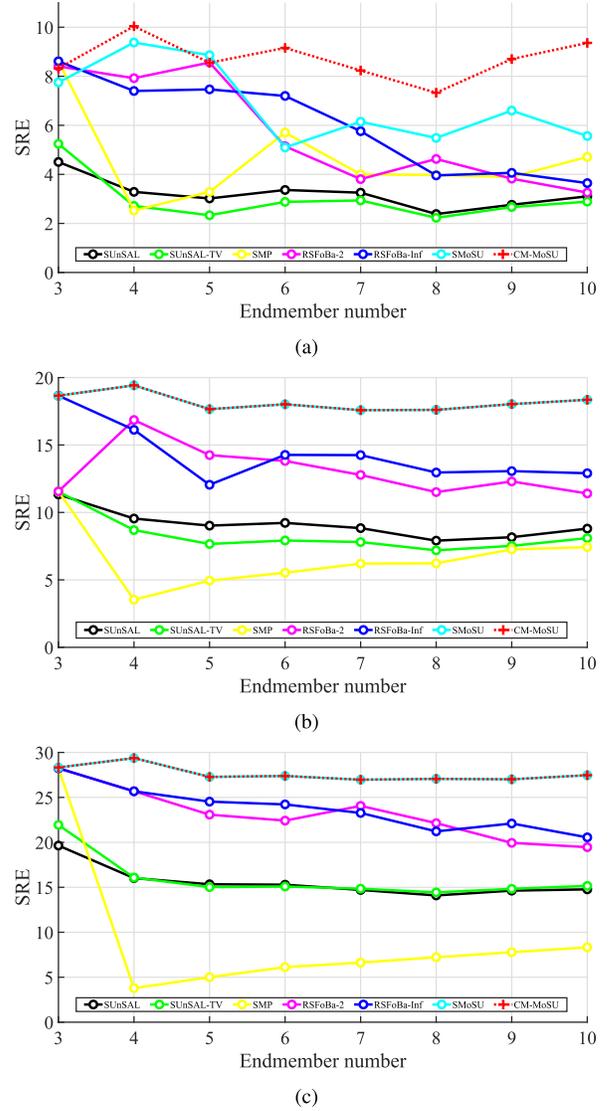


Fig. 4. Comparison of abundances obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, and CM-MoSU on synthetic data 1. The number of active endmembers is varying from 3 to 10. (a)–(c) Results for 20/30/40 dB data, respectively.

RSFoBa-Inf and SMoSU are 1 for 20-dB data. These three algorithms could find all the active spectra from the library. However, in general, CM-MoSU and SMoSU bring smaller FPRs compared with RSFoBa-Inf. In other words, they both obtain less redundant endmembers. Spectrally, CM-MoSU and SMoSU find the active spectra exactly without any redundancy for 30- and 40-dB data. It can be concluded that multi-objective optimization is effective for sparse unmixing.

To further compare the performance of endmember extraction process by SMoSU and CM-MoSU, we propose a new metric, residual reconstruction error (RRE), which is defined by

$$\text{RRE} \equiv \|\mathbf{A}_r \mathbf{X}_r\|_F^2 \quad (29)$$

where  $\mathbf{A}_r = \mathbf{A}_s / \mathbf{A}_{s^*}$  is the matrix containing all the residual endmembers, and  $\mathbf{X}_r$  is the corresponding abundance matrix. The physical meaning of RRE is that if the residual endmembers have occupied less reconstruction value, they will present

TABLE II

TPR AND FPR RESULTS OF SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, AND CM-MoSU ON SYNTHETIC DATA 2 WITH 30-dB CORRELATED NOISE

	$k$	3	4	5	6	7	8	9	10
SMP	TPR	1	7.50e-1	8.00e-1	8.33e-1	8.57e-1	8.75e-1	8.89e-1	9.00e-1
	FPR	2.00e-3	6.10e-3	6.10e-3	3.66e-2	8.15e-2	5.10e-2	1.53e-1	4.10e-2
RSFoBa-2	TPR	1	1	1	1	1	1	1	1
	FPR	2.00e-3	4.00e-3	6.10e-3	3.66e-2	9.37e-2	8.37e-2	1.35e-1	1.31e-1
RSFoBa-Inf	TPR	1	1	1	1	1	1	1	1
	FPR	1.00e-3	6.10e-3	6.10e-3	2.85e-2	4.07e-2	1.12e-1	1.06e-1	1.56e-1
SMoSU	TPR	1	1	1	1	1	1	1	1
	FPR	0	0	0	0	0	0	0	0
CM-MoSU	TPR	1	1	1	1	1	1	1	1
	FPR	0	0	0	0	0	0	0	0

TABLE III

TPR AND FPR RESULTS OF SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, AND CM-MoSU ON SYNTHETIC DATA 2 WITH 40-dB CORRELATED NOISE

	$k$	3	4	5	6	7	8	9	10
SMP	TPR	1	1	8.00e-1	8.33e-1	8.57e-1	8.75e-1	8.89e-1	9.00e-1
	FPR	2.00e-3	4.00e-3	6.10e-3	3.86e-2	8.55e-2	4.29e-2	1.51e-1	4.30e-2
RSFoBa-2	TPR	1	1	1	1	1	1	1	1
	FPR	2.00e-3	6.10e-2	2.00e-3	3.66e-2	3.26e-2	8.16e-2	1.27e-1	1.54e-1
RSFoBa-Inf	TPR	1	1	1	1	1	1	1	1
	FPR	2.00e-3	6.10e-3	6.10e-3	3.05e-2	7.33e-2	1.00e-1	1.19e-1	1.49e-2
SMoSU	TPR	1	1	1	1	1	1	1	1
	FPR	0	0	0	0	0	0	0	0
CM-MoSU	TPR	1	1	1	1	1	1	1	1
	FPR	0	0	0	0	0	0	0	0

TABLE IV

RRE COMPARISON BY S-MoSU AND CM-MoSU ON SYNTHETIC DATA 1 WITH 20-dB NOISE

	Endmember Number							
	3	4	5	6	7	8	9	10
SMoSU	47.01	44.13	40.00	34.25	23.56	26.78	28.99	<b>20.06</b>
CM-MoSU	<b>29.73</b>	<b>19.35</b>	<b>12.14</b>	<b>17.24</b>	<b>17.84</b>	<b>24.53</b>	<b>24.24</b>	22.11

less influence on the abundance inversion results. In other words, if RRE is smaller, the hyperspectral image will be more likely to be reconstructed by  $\mathbf{A}_{S^*}$ . Therefore, smaller RRE corresponds to better results.

Table IV displays the RRE values by SMoSU and CM-MoSU on synthetic data 1 with 20-dB noise. We can find that in most cases the RRE by CM-MoSU is smaller than SMoSU. Generally, RRE declines with the increase in real endmember numbers. This is not difficult to explain. Here, we take an example. An algorithm may have selected five endmembers in the case of three real endmembers, where two of them are residual. This algorithm may select

12 endmembers in the case of 10 real endmembers, where there are also 2 residual ones. Obviously, the former tends to present higher  $\mathbf{X}_r$  which may result in higher RRE. Compared with SMoSU, CM-MoSU has shown lower RRE values. According to (6), the reconstruction error using the endmembers extracted by CM-MoSU is more similar to  $(\mathbf{Y} - \mathbf{A}_{S^*}\mathbf{X}_{S^*})$ , and thus the abundance  $\mathbf{X}_{S^*}$  will be more accurate. Note that when endmember number is 10, SMoSU presents lower RRE than CM-MoSU, but its abundance inversion results are slightly worse. This is when endmember number is 10, the residual endmembers will occupy very few proportions which has little influence on the abundance results.

TABLE V  
TPR AND FPR RESULTS OF SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, AND CM-MoSU ON SYNTHETIC DATA 2

SNR	ROC	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU	CM-MoSU
20dB	TPR	1	1	1	1	1
	FPR	8.10e-3	8.10e-3	1.83e-2	4.10e-3	4.10e-3
30dB	TPR	1	1	1	1	1
	FPR	6.10e-3	8.10e-3	1.01e-2	2.00e-3	2.00e-3
40dB	TPR	1	1	1	1	1
	FPR	6.10e-3	8.10e-3	8.10e-3	0	0

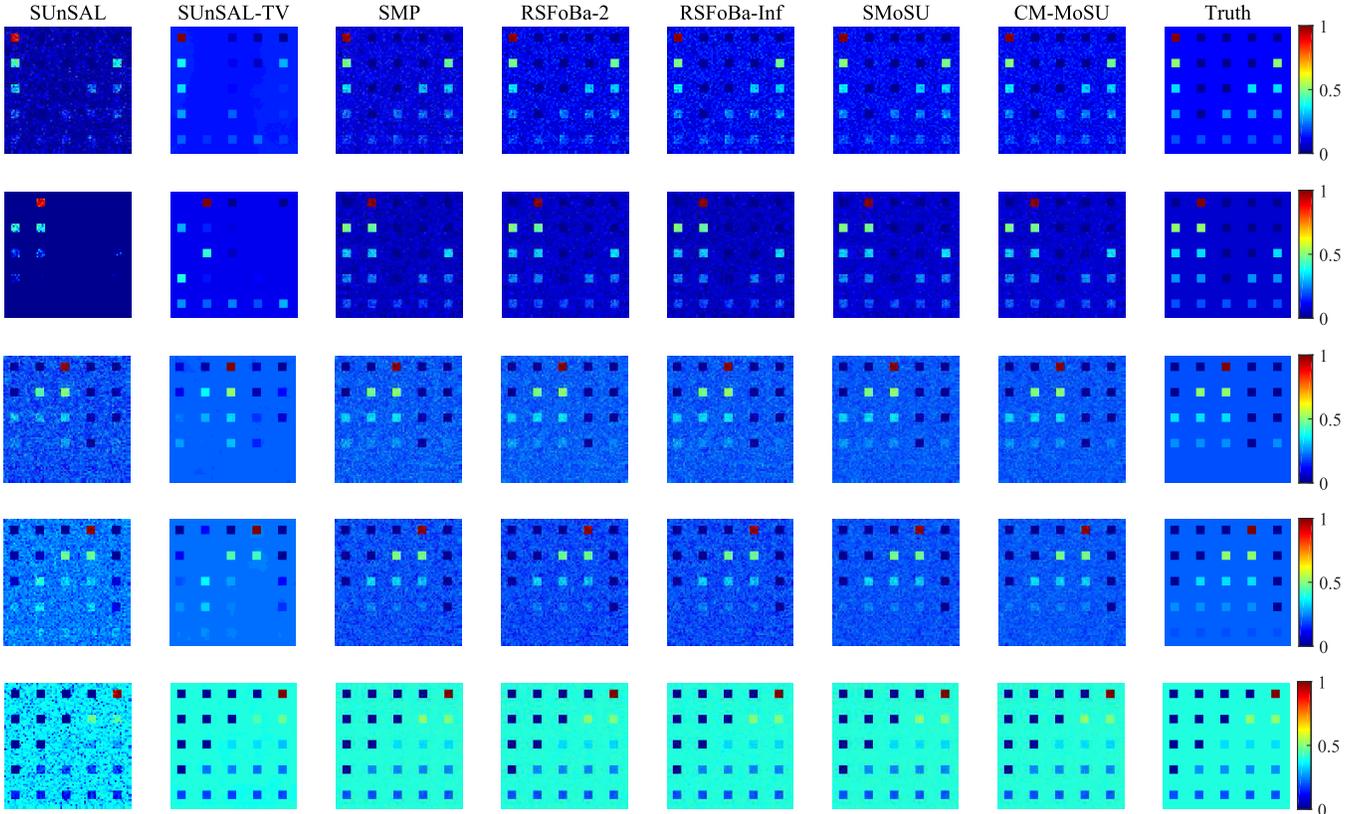


Fig. 5. Comparison of abundance maps on synthetic data 2 with 30-dB correlated noise. (Left to right) Abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, CM-MoSU, and the truth. (Top to bottom) Maps corresponding to endmembers #1 to #5.

TABLE VI  
SRE RESULTS OF SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, AND CM-MoSU ON SYNTHETIC DATA 2

SNR	SUnSAL	SUnSAL-TV	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU	CM-MoSU
20dB	4.5568	10.6189	5.2507	8.5797	5.5505	9.1531	<b>10.7630</b>
30dB	8.5833	15.1405	15.0461	15.9768	15.6683	18.7173	<b>19.1302</b>
40dB	12.9890	24.5638	24.6527	25.7949	25.7949	<b>29.3208</b>	<b>29.3208</b>

3) *Abundance Assessment*: Fig. 4 shows the results of SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, and CM-MoSU with different endmembers number. The unmixing accuracy is better for weaker noise cases in general. In Fig. 4(a), when the noise is strong (20 dB), CM-MoSU surpasses the other six methods in most cases

( $\hat{k} = 4, 6-10$ ), especially when  $\hat{k}$  is large. Compared with SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, and RSFoBa-Inf, CM-MoSU shows more obvious advantage than SMOsU with the increasing of  $\hat{k}$ . These results verify that our improvement is effective. From Fig. 4(b) and (c), the trends of all algorithms are similar for 30- and 40-dB data. CM-MoSU and SMOsU

TABLE VII  
TPR AND FPR RESULTS OF SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, AND CM-MoSU ON SYNTHETIC DATA 3

SNR	ROC	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU	CM-MoSU
20dB	TPR	1	1	1	1	1
	FPR	6.10e-3	6.95e-2	8.20e-3	1.64e-2	1.64e-2
30dB	TPR	1	1	1	1	1
	FPR	1.43e-2	7.16e-2	4.10e-3	4.10e-3	4.10e-3
40dB	TPR	1	1	1	1	1
	FPR	7.98e-2	6.95e-2	1.23e-2	0	0

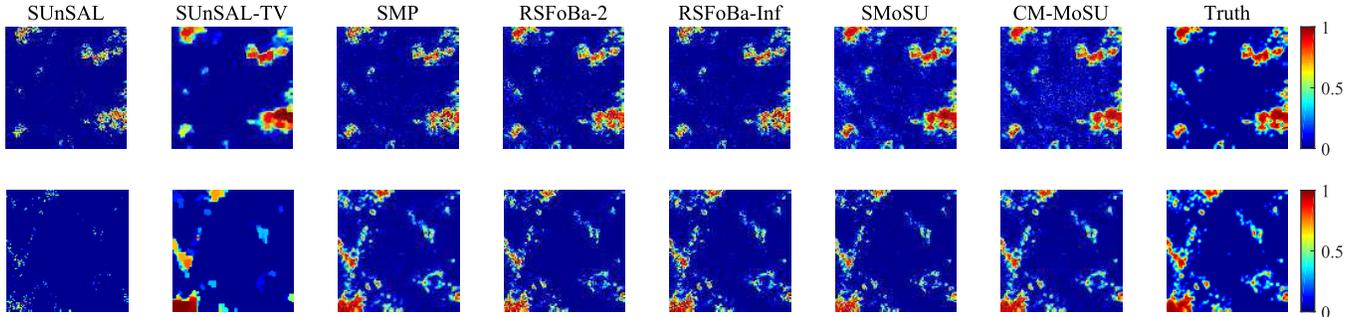


Fig. 6. Abundance maps of endmember #1 and #3 on synthetic data 3 with 20-dB correlated noise. (Left to right) Abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, CM-MoSU, and the truth, respectively. (Top to bottom) Maps corresponding to endmember #1 and #3, respectively.

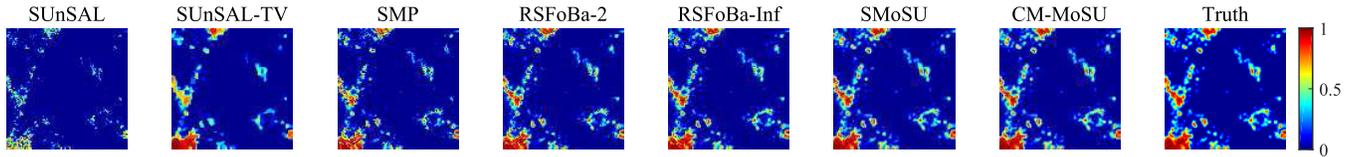


Fig. 7. Abundance maps of endmember #3 on synthetic data 3 with 30-dB correlated noise. (Left to right) Abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, CM-MoSU, and the truth, respectively.

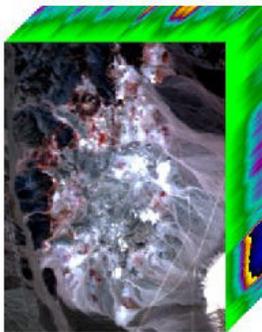


Fig. 8. Subscene of Cuprite data set.

are both better than the other five algorithms in these cases. They obtain the same results here, since they both find the true endmembers exactly.

### B. Synthetic Data 2

Synthetic data 2 is obtained using the data generation approach introduced by [25]. It has  $75 \times 75$  pixels and 5 endmembers. The 20-/30-/40-dB correlated noise is also added to it, respectively. Table V presents the endmember selection results (TPR and FPR) of all the seven algorithms. From Table V, the selected endmembers of these algorithms all cover

the true active spectra, since their TPRs are all 1. However, the number of inactive endmembers is different in their results. CM-MoSU and SMoSU have the smallest FPR for all the cases and obtain the true endmembers without any redundancy for 40-dB data. The abundance estimation results of these algorithms are presented in Table VI. As shown in Table VI, CM-MoSU and SMoSU obtain the same best performance for 40-dB data, because they both find the true endmembers exactly. However, CM-MoSU surpasses SMoSU for 20- and 30-dB data. Although they obtain the same FPRs, the redundant endmembers found by CM-MoSU are more effective. This result indicates that CM-MoSU has some superiority in high noise data. Fig. 5 shows abundance maps of the five endmembers. From left to right, abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU, CM-MoSU, and the truth, respectively, are shown. From top to bottom, the maps corresponding to endmembers #1 to #5 are shown. According to Fig. 5, CM-MoSU presents closer maps to the truth.

### C. Synthetic Data 3

Synthetic data 3 is also generated based on [25], which contains  $100 \times 100$  pixels and nine endmembers. Table VII present the endmember selection results of all the seven

TABLE VIII  
SRE RESULTS OF SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, AND CM-MoSU ON SYNTHETIC DATA 3

SNR	SUnSAL	SUnSAL-TV	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU	CM-MoSU
20dB	3.5823	6.8356	11.7575	9.1401	10.4579	12.3412	<b>12.4388</b>
30dB	8.0323	11.8903	15.2202	16.5553	19.7419	22.2353	<b>23.0996</b>
40dB	12.9896	17.1311	21.6863	25.0255	27.1758	<b>33.8674</b>	<b>33.8674</b>

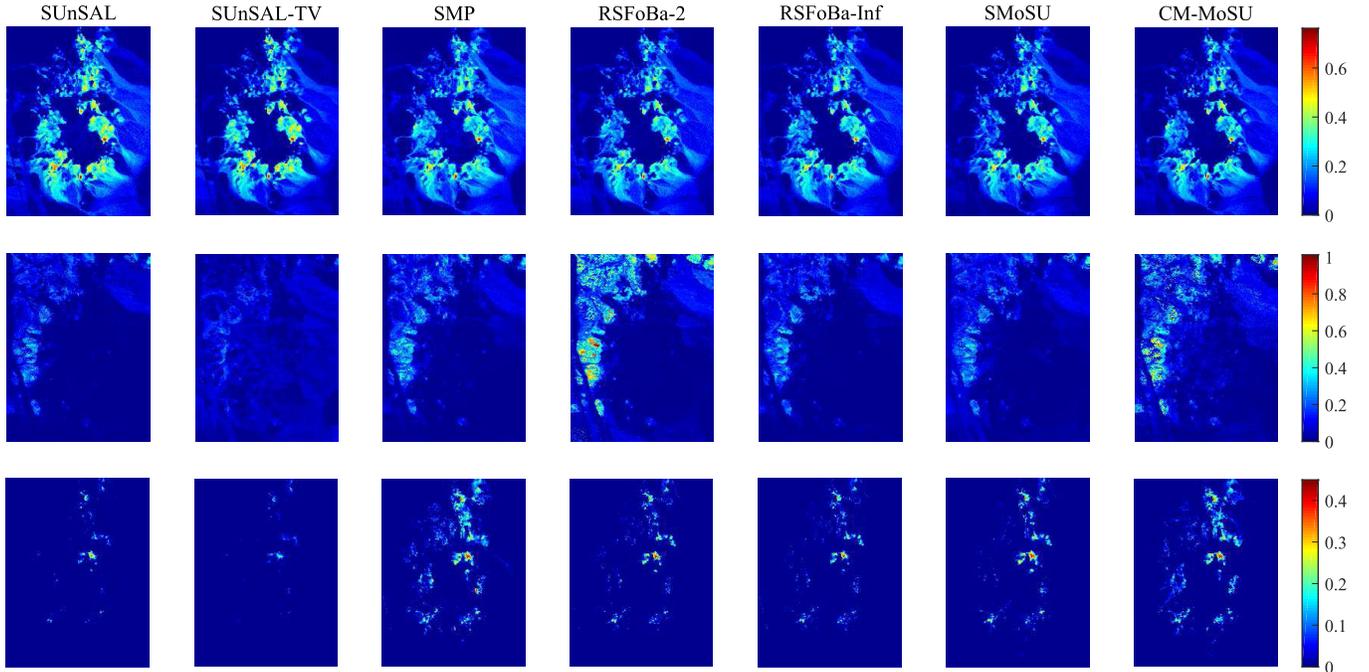


Fig. 9. Comparison of abundance maps on Cuprite data. (Left to right) Abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, and CM-MoSU, respectively. (Top to bottom) Maps corresponding to Alunite + Muscovite/Kaolinite, Hematite, Alunite.

algorithms. From Table VII, CM-MoSU and SMOsU find the true endmembers exactly in the 40-dB noise case and obtain the same FPR when the noise is 20 and 30 dB. The abundance estimation results are presented in Table VIII. From Table VIII, the advantage of CM-MoSU mainly reflects on the 20- and 30-dB data. The abundance maps of endmember #1 and #3 for 20-dB data and abundance maps of endmember #3 for 30-dB data are shown in Figs. 6 and 7, respectively. In general, the results of CM-MoSU are closer the truth compared with the others.

#### D. Real-World Experiments

For real-world experiments, a subscene of the well-known AVIRIS Cuprite data set<sup>2</sup> is used, as shown in Fig. 8. It contains  $250 \times 190$  pixels. The water absorption bands and some noisy bands (bands 1, 2, 221–224) are taken out from the image and spectral library. A qualitative comparison is presented for this data set, since the ground truth is unavailable. Fig. 9 presents a visible comparison of the abundance maps obtained by the seven algorithms. From left to right, abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMOsU, and CM-MoSU, respec-

tively, are shown. From top to bottom, the maps corresponding to Alunite+Muscovite/Kaolinite, Hematite, and Alunite are shown. From Fig. 9, the seven methods obtain similar results for Alunite+Muscovite/Kaolinite. CM-MoSU and RSFoBa-2 find the distribution of Hematite better than the other methods. The abundance maps of SMP and CM-MoSU have more clear outlines for Alunite.

#### V. CONCLUSION

In this paper, a classification-based model is designed to improve the performance of SMOsU, especially for strong noise data. The proposed CM-MoSU is under the framework of SMOsU. In CM-MoSU, CM is used to generate new individuals. The whole population is divided into two categories: individuals that have better performance are denoted to be positive and negative otherwise. CM takes advantages of the positive individuals, so that the optimization task could pay more attention to the high-quality feasible space. The information about negative individuals is used to produce diverse individuals. To further improve the individual diversity, the bitwise flipping strategy of SMOsU is retained with a probability in CM-MoSU. Furthermore, some theoretical analyses are presented for CM-MoSU. The failing probability

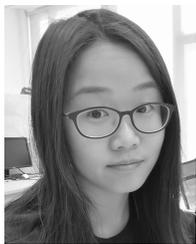
<sup>2</sup><http://lesun.weebly.com/hyperspectral-data-set.html>

of generating individuals from CM and the overall failing probability of CM-MoSU are quantitatively analyzed.

Synthetic and real-world experiments are used to test the performance of CM-MoSU. Six state-of-the-art sparse unmixing algorithms are compared with CM-MoSU, including SMOsU. The effect of positive size in CM-MoSU is analyzed. In addition, a new criterion is introduced to assess the endmember selection performance of subset selection-based sparse unmixing methods.

## REFERENCES

- [1] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018. doi: [10.1109/TGRS.2018.2828161](https://doi.org/10.1109/TGRS.2018.2828161).
- [2] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "CoinNet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 816–820, May 2019.
- [3] Q. Wang, Z. Meng, and X. Li, "Locality adaptive discriminant analysis for spectral-spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2077–2081, Nov. 2017.
- [4] X. Lu, W. Zhang, and X. Li, "A hybrid sparsity and distance-based discrimination detector for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 56, pp. 1704–1717, Mar. 2018.
- [5] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [6] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [7] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.
- [8] 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.
- [9] 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.
- [10] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [11] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, Oct. 1999.
- [12] C.-I. Chang, C.-C. Wu, W.-M. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.
- [13] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, Mar. 2012.
- [14] T. Uezato, R. J. Murphy, A. Melkumyan, and A. Chlingaryan, "A novel spectral unmixing method incorporating spectral variability within endmember classes," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2812–2831, May 2016.
- [15] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [16] W. Wang and Y. Qian, "Adaptive  $L_{1/2}$  sparsity-constrained NMF with half-thresholding algorithm for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 8, no. 6, pp. 2618–2631, Jun. 2015.
- [17] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016.
- [18] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Blind hyperspectral unmixing using total variation and  $\ell_q$  sparse regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6371–6384, Nov. 2016.
- [19] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
- [20] 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.
- [21] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using  $l_0$  norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5495–5508, Sep. 2018.
- [22] Z. Shi, W. Tang, Z. Duren, and Z. Jiang, "Subspace matching pursuit for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3256–3274, Jun. 2014.
- [23] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2010, pp. 1–4.
- [24] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.
- [25] 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, Nov. 2012.
- [26] R. Feng, Y. Zhong, and L. Zhang, "Adaptive non-local Euclidean medians sparse unmixing for hyperspectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 97, pp. 9–24, Nov. 2014.
- [27] S. Zhang, J. Li, K. Liu, C. Deng, L. Liu, and A. Plaza, "Hyperspectral unmixing based on local collaborative sparse regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 631–635, May 2016.
- [28] C. Shi and L. Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sens. Environ.*, vol. 149, pp. 70–87, Jun. 2014.
- [29] X. Fu, W.-K. Ma, J. M. Bioucas-Dias, and T.-H. Chan, "Semiblind hyperspectral unmixing in the presence of spectral library mismatches," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5171–5184, Sep. 2016.
- [30] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.
- [31] W. Tang, Z. Shi, and Y. Wu, "Regularized simultaneous forward-backward greedy algorithm for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5271–5288, Sep. 2014.
- [32] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 544–557, Jan. 2016.
- [33] B. Pan, Z. Shi, and X. Xu, "Analysis for the weakly Pareto optimum in multiobjective-based hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3729–3740, Jun. 2019.
- [34] A. Ma, Y. Zhong, and L. Zhang, "Adaptive multiobjective memetic fuzzy clustering algorithm for remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4202–4217, Aug. 2015.
- [35] M. Gong, H. Li, E. Luo, J. Liu, and J. Liu, "A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 234–248, Apr. 2017.
- [36] X. Jiang, M. Gong, H. Li, M. Zhang, and J. Li, "A two-phase multiobjective sparse unmixing approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 508–523, Jan. 2018.
- [37] X. Xu and Z. Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 54–69, Feb. 2017.
- [38] X. Xu, Z. Shi, and B. Pan, " $\ell_0$ -based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 46–58, Jul. 2018.
- [39] J. Ceberio, E. Irurozki, A. Mendiburu, and J. A. Lozano, "A review on estimation of distribution algorithms in permutation-based combinatorial optimization problems," *Prog. Artif. Intell.*, vol. 1, no. 1, pp. 103–117, 2012.
- [40] M. Zangari, A. Mendiburu, R. Santana, and A. Pozo, "Multiobjective decomposition-based Mallows Models estimation of distribution algorithm. A case of study for permutation flowshop scheduling problem," *Inf. Sci.*, vols. 397–398, pp. 137–154, Aug. 2017.
- [41] M. Z. de Souza, R. Santana, A. T. R. Pozo, and A. Mendiburu, "MOEA/D-GM: Using probabilistic graphical models in MOEA/D for solving combinatorial optimization problems," Nov. 2015, *arXiv:1511.05625*. [Online]. Available: <https://arxiv.org/abs/1511.05625>
- [42] Y. Yu, H. Qian, and Y.-Q. Hu, "Derivative-free optimization via classification," in *Proc. 30th AAAI Conf. Artif. Intell. (AAAI)*, 2016, pp. 2286–2292.
- [43] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.



**Xia Xu** received the B.S. and M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015, respectively. She is currently pursuing the Ph.D. degree with the Image Processing Center, School of Astronautics, Beihang University, Beijing, China.

Her research interests include hyperspectral unmixing and multi-objective optimization.

Dr. Xu serves as a reviewer for several journals including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE

TRANSACTIONS ON CYBERNETICS.



**Bin Pan** received the B.S. degree from the School of Astronautics, Beihang University, Beijing, China, in 2013, where he is currently pursuing the Ph.D. degree with the Image Processing Center.

His research interests include machine learning and image processing.

Mr. Pan serves as a reviewer for several journals, including *ISPRS Journal of Photogrammetry and Remote Sensing* and *IEEE Transactions on Geoscience and Remote Sensing*.



**Zhenwei Shi** (M'13) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005.

From 2005 to 2007, he was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang

University, Beijing. He has authored or coauthored more than 100 scientific papers in related journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi received the Best Reviewer Awards for his service to IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) in 2017. He has been an Associate Editor for the *Infrared Physics and Technology* since 2016.

**Xuelong Li** (M'02–SM'07–F'12) is currently a Full Professor with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China.