ℓ_0-based Sparse Hyperspectral Unmixing Using Spectral Information and a Multi-Objectives Formulation ☆

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

[a]*Image Processing Center, School of Astronautics, Beihang University, Beijing, PR China*
[b]*State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, PR China*
[c]*Beijing Key Laboratory of Digital Media, Beihang University, Beijing, PR China*

Abstract

Sparse unmixing aims at recovering pure materials from hyperpspectral images and estimating their abundance fractions. Sparse unmixing is actually ℓ_0 problem which is NP-hard, and a relaxation is often used. In this paper, we attempt to deal with ℓ_0 problem directly via a multi-objective based method, which is a non-convex manner. The characteristics of hyperspectral images are integrated into the proposed method, which leads to a new spectra and multi-objective based sparse unmixing method (SMoSU). In order to solve the ℓ_0 norm optimization problem, the spectral library is encoded in a binary vector, and a bit-wise flipping strategy is used to generate new individuals in the evolution process. However, a multi-objective method usually produces a number of non-dominated solutions, while sparse unmixing requires a single solution. How to make the final decision for sparse unmixing is challenging. To handle this problem, we integrate the spectral characteristic of hyperspectral images into SMoSU. By considering the spectral correlation in hyperspectral data, we improve the Tchebycheff decomposition function in SMoSU via a new regularization item. This regularization item is able to enforce the individual divergence in the evolution process of SMoSU. In this way, the diversity and convergence of population is further balanced, which is beneficial to the concentration of individuals. In the experiments part, three synthetic datasets and one real-world data are used to analyse the effectiveness of SMoSU, and several state-of-art sparse unmixing algorithms are compared.

1. Introduction

Benefit from the development of remote sensing technique, imagery spectral resolution has been improved significantly and hyperspectral observation capability is formed. Hyperspectral images usually contain hundreds of spectrum bands, covering visible to thermal-infrared regions (Ma et al. 2014; Zhong et al. 2018). The abundant spectral information of hyperspectral image contributes to many practical applications, such as environmental monitoring and geological exploration (Ruitenbeek et al. 2006; Pan et al. 2017a). However, the spatial resolution of hyperspectral images is usually low and different features are always homogeneously mixed (Willett et al. 2014; Wang et al. 2017). Thus a single pixel always contains more than one land cover types, resulting in mixed pixels. The complex mixing of different features brings great challenge to hyperpsectral image processing (Pan et al. 2016; Wang et al. 2016; Zhou and Wei 2016; Pan et al. 2017b). Unmixing aims at recovering pure materials spectra (*endmembers*) of a hyperspectral image, as well as their corresponding fractions (*abundances*). The abundance values are proportions representing the percentage of endmembers in a pixel region (Keshava and Mustard 2002). Accordingly, hyperspectral unmixing is generally processed under two steps: (i) identifying the endmembers, (ii) quantifying the abundance fractions (Bioucas-Dias et al. 2012; Zhong et al. 2016).

Sparse regression based unmixing is a hot topic in recent years, which does not need to assume pure materials in hyperspectral images. Due to the simplicity and flexibility, linear mixing model (LMM) is the most widely used (Bioucas-Dias et al. 2012; Heylen et al. 2014; Zhu et al. 2014b). LMM characterizes the mixture without considering the effects of multiple scattering and intimate mixture. In sparse unmixing, the mixed image data is represented by pure signatures from a spectral library which is known in advance (Iordache et al. 2011). In this way the estimation of abundances is no longer dependent on the presence of pure pixels. Sparse unmixing aims at determining an optimal subset of the pure materials, which may be a very small proportion relative to the library. Mathematically speaking, it is a ℓ_0 norm-based combinational problem, which is non-convex and NP-hard. The most commonly used approach to solve this problem is approximating it to a convex ℓ_1 norm regularized optimization problem (Bioucas-Dias and Figueiredo 2010; Iordache et al. 2014a). To further approximate the ℓ_0 norm, $\ell_p(0 < p < 1)$ norm based methods are developed (Qian et al. 2011; Zhu et al. 2014a). In order to obtain a better unmixing performance, quite a few studies

consider exploring the characteristics of hyperspectral image and the spectral library, such as the contextual information among pixels (Feng et al. 2014, 2016), the signature differences between hyperspectral image and spectral library (Shi and Wang 2014; Zare and Ho 2014). However, ℓ_1 or ℓ_p norm is only a relaxation of the original ℓ_0 problem. Using such a relaxation may lead to errors in the unmixing results. Another type of works try to deal with the ℓ_0 norm problem directly using greedy algorithms. In (Shi et al. 2014), a subspace matching pursuit sparse unmixing method was proposed. Considering the high correlation of spectral library, the greedy selection is conducted based on the whole hyperspectral image data. In (Tang et al. 2014), the forward greedy step and the backward greedy step was combined to provide a more stable selection and less probabilities of local optima. Although greedy-based methods need no approximation, there are many sensitive parameters which have to adjust manually. Moreover, these methods usually encounter problems of endmember missing and redundancy (Shi et al. 2014).

Recently, multi-objective optimization (MO) has been proposed to solve ℓ_0 norm-based problems directly without any smoothing (Deb and Jain 2014; Jain and Deb 2014; Ma et al. 2015; Sun et al. 2016; Zhang and Tao 2017; Ma et al. 2018). The major target of MO is to find a non-dominated solution set which provides a good trade-off for objective functions (Miettinen 1999; Deb and Kalyanmoy 2001). MO has made great progress in combinatorial optimization (Li et al. 2014)(Xue et al. 2016). Researchers have verified that MO can provide good solutions for many NP-hard problems, where it can recover the best-so-far guaranteed approximate solution within limited iterations. In Yu et al. (2013), an isolation-based MO framework was proposed, which could achieve the best-achievable result on minimum k-set cover problem with an H_k-approximation ratio. In the next few years, Qian *et al.* further explored the evolutionary problem and provided many theoretical supports on some NP-hard problems. In Qian et al. (2015a), Pareto optimization, penalty function method and greedy algorithms were compared theoretically on the minimum cost coverage problem. Pareto optimization was proved more efficient than penalty function method. It was also found to be positive on a special case of the problem, when compared with greedy algorithm. In Qian et al. (2015b), Pareto optimization based subset selection (POSS) was found that it was able to achieve the best-so-far approximation guarantee obtained by greedy algorithms on sparse regression. Later, POSS is sped up through paralleling (Qian et al. 2016).

For sparse unmixing, the balance of reconstruction error and endmember sparsity is precisely in line with the goal of MO. However, different from single-objective optimization, a number of

conflicting objective functions need to be optimized simultaneously in MO (Deb and Kalyanmoy 2001). In other words, it is difficult to select a specific solution that is optimal to the two objective functions. Therefore, MO is expected to determine a set of non-dominated solutions as close as the Pareto-optimal front to give a trade-off among objectives (Zhou et al. 2011). In practical applications such as sparse unmixing, there must be a specific strategy to determine the unique solution.

There are few works that try to handle the sparse unmixing problem by MO. In Xu and Shi (2017), sparse unmixing was transformed into a bi-objective optimization problem using POSS and non-dominated sorting genetic algorithm-II (NSGA-II). This method was a semi-automatic manner which requires selecting an individual from the final solution set manually. In Gong et al. (2017), considering the large scale of spectral library and heavy computing load caused by high dimensional problem, the spectral library was grouped and a cooperation strategy was designed among the groups. This algorithm optimized the abundance matrices and used the knee point in the Pareto front as the final solution. However, although these methods have presented good performance, how to determine the final solution from the obtained non-dominated front is still challenging for sparse unmixing. If suboptimal point is picked as the final solution, the endmember missing or redundancy is likely to happen. In this case, the unmixing accuracy is suffered significantly. Furthermore, endmember sparsity in most existing methods relies heavily on the parameter settings of the mutation and crossover operator, which may result in an inadequate sparsity.

In this paper, a new multi-objective optimization based sparse unmixing method is proposed, which takes full advantage of the spectral characteristic in hyperspectral images. We term the proposed work as integrating spectra and multi-objective for ℓ_0 sparse unmixing (SMoSU). SMoSU is developed under the framework of the multi-objective evolutionary algorithm based on decomposition (MOEA/D) (Zhang and Li 2007). MOEA/D could provide a non-Pareto criterion, which leads to a fast evolutionary speed and low computational complexity. In SMoSU, sparse unmixing is transformed to a bi-objective discrete optimization problem, where reconstruction error and endmember sparsity error are taken as the two conflicting objectives. Inspired by the binary coding of minimum k-set cover problem in (Yu et al. 2013), we encode the spectral library in a binary vector and use a bit-wise flipping strategy for individual generation. Each bit of the binary representation indicates the corresponding spectrum. The probability-based bit-wise flipping is verified to be effective in global search and does not need to adjust many parameters (Yu et al. 2013; Qian

4

et al. 2015b). However, as is discussed above, a set of non-dominated solutions are generated by MOEA/D, which is hard to select a final solution. In this paper, we improve the Tchebycheff decomposition approach in MOEA/D by integrating the spectral characteristic of hyperspectral data. The improved decomposition function takes spectral correlations among individuals into account. A regularization item that enforces the individual divergence is included in SMoSU. By this means the diversity and convergence of population are further balanced, which is beneficial to the concentration of individuals. The major contributions of SMoSU can be summarized as follows:

- We introduce a new multi-objective based sparse unmixing method which could solve the ℓ_0 non-convex optimization without any relaxation. The reconstruction error and endmember sparsity error are considered as two objectives in SMoSU.

- We use a binary code strategy for spectral signatures in the library and proposed a bit-wise flipping approach for individual generation based on the framework of MOEA/D. In this case, the optimal subset selection of spectral library is transformed to finding an optimal binary vector.

- To overcome the difficulty of selecting the final solution, we integrate the spectral characteristic of hyperspectral images into the multi-objective framework. The Tchebycheff decomposition approach is regularized by the spectral correlations among different individuals, so as to further balance the population divergence and convergence.

The rest of this paper is organized as follows. Section 2 introduces the background about sparse unmixing and MOEA/D. In Section 3, detailed description about SMoSU is presented. Experimental results and analysis are shown in Section 4. We conclude this paper in Section 5.

2. Backgrounds

In this section, we first give a brief introduction about the model and optimization problem of linear sparse unmixing, then some basic concepts about multi-objective optimization is presented.

2.1. Linear sparse unmixing

In LMM, the spectra of mixed pixel is weighted linearly by several pure spectra in the library. Let $\mathbf{y} \in \mathbb{R}^{L \times 1}$ be the measured spectrum of a mixed pixel with L bands and $\mathbf{A} \in \mathbb{R}^{L \times m}$ be the

5

spectral library with m pure materials. The linear model of each pixel can be written as

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

where $\mathbf{x} \in \mathbb{R}^{m \times 1}$ is the abundance fractions of the mixed pixel, $x_i \in \mathbf{x}$ stands for the proportion of the ith endmember, $\mathbf{n} \in \mathbb{R}^{L \times 1}$ is the error term. Considering the abundance fractions geometrically, the following constrains need to be considered

$$
\begin{aligned}
&\text{ANC} : x_i \geq 0 \\
&\text{ASC} : \sum_{i=1}^{m} x_i = 1
\end{aligned}
\tag{2}
$$

where ANC is short for the abundance nonnegativity constraint and constrains nonnegative areal presences to every endmember; ASC is short for the abundance sum constraint and expects a complete decomposition to the observed spectral signature of each pixel. For the whole hyperspectral image with n pixels, the linear model can be written as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{3}$$

where $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_n] \in \mathbb{R}^{L \times n}$ is the image data matrix, $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ is the abundance matrix of all the n pixels, $\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_n] \in \mathbb{R}^{L \times n}$ is the corresponding error term.

In view of the fact that all pixels in a hyperspectral image share a small number of pure materials relative to the scale of spectral library, there should be only a few nonzero lines for the abundance matrix $\mathbf{X}$ (Iordache et al. 2014a). Thus the sparse unmixing problem can be described as

$$
\begin{aligned}
(P_0) : &\min_{\mathbf{X}} \|\mathbf{X}\|_{row-0} \\
&s.t. \ \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\mathrm{F}} \leq \delta, \mathbf{X} \geq 0
\end{aligned}
\tag{4}
$$

where $\|\mathbf{X}\|_{row-0}$ is the number of nonzero rows in $\|\mathbf{X}\|$, indicating the estimated number of endmembers, $\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\mathrm{F}}$ is the Frobenius norm of $\mathbf{Y} - \mathbf{A}\mathbf{X}$, δ is the error tolerance. It has been studied that ANC gives an automatic imposition to a generalized ASC (Iordache et al. 2011), so the ASC is not added in Eq. (4).

2.2. Mathematical description of MO

MOs need to optimize several conflicting objectives simultaneously in practice. Suppose there exist N objectives and m decision variables. MO can be described mathematically as

$$\min F(\mathbf{u}) = [f_1(\mathbf{u}), f_2(\mathbf{u}), ..., f_N(\mathbf{u})]^{\mathrm{T}}$$
$$s.t. \ h_i(\mathbf{u}) \leq 0, i = 1, ..., v_1$$
$$h_j(\mathbf{u}) = 0, j = v_1 + 1, ..., v_2 \tag{5}$$
$$\mathbf{u} \in \Omega \subseteq \mathbb{R}^m$$

where $\mathbf{u} = (u_1, ..., u_m)$ denotes the m decision variables, Ω is the decision space, $F : \Omega \rightarrow \mathbb{R}^N$ contains the objective functions, $h_i(\mathbf{u})$ and $h_j(\mathbf{u})$ are inequality and equality constrains of the MO problem, v_1 and v_2 are their numbers respectively. On this basis, the following main definitions are given (Deb and Kalyanmoy 2001):

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

$$\forall i \in \{1, 2, ..., N\} : f_i(\mathbf{u}^1) \leq f_i(\mathbf{u}^2) \wedge \exists j \in \{1, 2, ..., N\} : f_j(\mathbf{u}^1) < f_j(\mathbf{u}^2) \tag{6}$$

Definition 2 (Pareto Optimality): A decision vector $\mathbf{u}^* \in \Omega$ is a Pareto optimal solution of (5) if and only if $\nexists \mathbf{u} \in \Omega, \mathbf{u} \prec \mathbf{u}^*$.

Definition 3 (Pareto Set): Pareto set is the collection of all the Pareto optimal solutions: $\{\mathbf{u}^* \mid \nexists \mathbf{u} \in \Omega, \mathbf{u} \prec \mathbf{u}^*\}$.

Definition 4 (Pareto Front): Pareto front is the corresponding objective values of Pareto Set: $\{F(\mathbf{u}^*) \mid \mathbf{u}^* \in \text{Pareto Set}\}$.

Some other definitions of multi-objective optimization are shown in Table 1.

Table 1: Some related definitions of multi-objective optimization.

individual	a solution which corresponds to a series of selected spectra from the library
population	a solution set that contains several individuals
population size	the number of individuals in a population
Pareto criterion	Pareto Dominance based evaluation criteria to judge the better individuals

The objective functions in MO are conflicting, which means that the reduction of an objective function value is at the expense of other's increases. It is hard to find a solution optimal to all

the objectives simultaneously. Thus MOs are expected to find a Pareto Set to bring a balance to all the objectives. Originally, MOs were solved by transforming to single-objective optimization problems. Recently, multi-objective evolutionary algorithms (MOEA) were proposed, followed by Pareto criterion based methods. MOEAs use a population based optimization mechanism. In each iteration, a number of candidate solutions are updated in parallel by a serial of operations such as crossover, mutation, selection, and elite-conservation.

In Zhang and Li (2007), Zhang *et al.* proposed an MOEA/D framework, which has been verified effective in several fields. Unlike traditional Pareto criterion based algorithms, MOEA/D is a mathematical programming based method that decomposes the MO problem to a set of single-objective optimization subproblems. These subproblems could provide a non-Pareto criterion without complex non-dominated sorting operations, which leads to a faster evolutionary speed for MOEA/D. Moreover, the computational complexity is low in MOEA/D, because it does not need complex methods to maintain population diversity (such as crowded distance methods).

3. SMoSU for sparse unmixing

Due to the non-convex and NP-hard of the original ℓ_0 norm problem, few works target at the (P_0) directly. Usually, penalty functions are used in most sparse unmixing studies. For example, using $\ell_q(0 < q \leq 1)$ norm for relaxation:

$$(P_1): \min_{\mathbf{X} \geq 0} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\mathrm{F}} + \lambda_1 \|\mathbf{X}\|_q + \mathcal{G}(\mathbf{Y}, \mathbf{X}, \lambda_i, i = 2, ..., r) \tag{7}$$

where $\|\mathbf{X}\|_q$ is the relaxation of $\|\mathbf{X}\|_{row-0}$, $\mathcal{G}(\mathbf{Y}, \mathbf{X}, \lambda_i, i = 2, ..., r)$ is a collection of the considered hyperspectral image characteristics, such as the spatial-contextual and low-rank constrains of abundances, and $\lambda_i(i = 1, ..., r)$ are regularization coefficients of the constrains and r is their number. If the optimization problem (P_1) is convex, it could be well solved based on alternating direction method of multipliers (Bioucas-Dias et al. 2012) or iterative thresholding method (Gong et al. 2013), but these regularization coefficients are sensitive and have an great effect on unmixing results. Greedy algorithms could avoid the approximation of ℓ_0 norm, but they are also sensitive to a number of parameters. Additionally, greedy methods are likely to face the problem of endmember missing or redundancy especially when the number of actual endmembers is large and the noise is strong (Shi et al. 2014).

Here, we consider the sparse unmixing problem as two conflicting subproblems: minimizing the reconstruction error $\|\mathbf{Y} - \mathbf{AX}\|_{\mathrm{F}}$ and adjusting the endmember sparsity $\|\mathbf{X}\|_{row-0}$. These functions are regarded as two objectives in MOEA. The obtained optimization problem is

$$\min_{\mathbf{X}} F(\mathbf{X}) = [f_1(\mathbf{X}), f_2(\mathbf{X})]^{\mathrm{T}}$$

$$f_1(\mathbf{X}) = \begin{cases} +\infty, & \|\mathbf{X}\|_{row-0} = 0 \ or \ \geq 2k \\ \|\mathbf{Y} - \mathbf{AX}\|_{\mathrm{F}}, & otherwise \end{cases} \tag{8}$$

$$f_2(\mathbf{X}) = \big|\|\mathbf{X}\|_{row-0} - k\big|$$

where k is the real endmember number. Because $f_2(\cdot)$ considered as an objective function rather than a regularizer, here we use the sparsity error instead the ℓ_0 sparsity to represent $f_2(\cdot)$. By this means the optimal value of $f_2(\cdot)$ is forced to be 0. A possible problem for Eq. (8) is that its solution space is relatively large. If it is directly handled, it is difficult to find the real optimal solution. In SMoSU we use a two-step approach to handle the problem of sparse unmixing: endmember selection and abundance estimation. We considered transforming the endmember selection to a binary code ($\{0, 1\}$) based optimization problem and inverse the abundance matrix $\mathbf{X}$ by using nonnegative least squares. In this case, the solution set would be shrunken a lot, and it may be more likely to find the real optimal solution. The optimization problem for endmember selection is:

$$\min_{\mathbf{s}} F(\mathbf{s}) = [f_1(\mathbf{s}), f_2(\mathbf{s})]^{\mathrm{T}}$$

$$f_1(\mathbf{s}) = \begin{cases} +\infty, & \|\mathbf{X}\|_{row-0} = 0 \ or \ \geq 2k \\ \|\mathbf{Y} - (\mathbf{AS}_m)\mathbf{X}\|_{\mathrm{F}}, & otherwise \end{cases} \tag{9}$$

$$f_2(\mathbf{s}) = \big|\|\mathbf{S}_m\|_1 - k\big|, \mathbf{S}_m = diag(\mathbf{s})$$

$f_1(\mathbf{s})$ is the reconstruction error function, $f_2(\mathbf{s})$ is the endmember sparsity error, $\mathbf{s} \in \mathbb{R}^m$ denotes a binary solution vector where the locations corresponding to the selected endmembers are set as "1" and "0" otherwise, as shown in Fig. 1. We transform Eq. (9) to the following form which is more conveniently to solve:

$$(P_{mo}) : \min_{\mathbf{s} \in \{0,1\}^m} F(\mathbf{s}) = [f_1(\mathbf{s}), f_2(\mathbf{s})]^{\mathrm{T}}$$

$$f_1(\mathbf{s}) = \begin{cases} +\infty, & \|\mathbf{s}\|_1 = 0 \ or \ \geq 2k \\ \|\mathbf{Y} - \mathbf{A_s X_s}\|_{\mathrm{F}}, & otherwise \end{cases} \tag{10}$$

$$f_2(\mathbf{s}) = \big|\|\mathbf{s}\|_1 - k\big|$$

9

$\|\mathbf{s}\|_1$ denotes the endmember number represented by a certain individual and $\mathbf{A_s}$ is the corresponding subset of spectral library, $\mathbf{X_s}$ is the abundances of endmembers. $\mathbf{s}$ and $\mathbf{X}$ are calculated alternatingly, but only $\mathbf{s}$ is the variable required optimized. When $\mathbf{s}$ is determined, $\mathbf{X}_s$ could be directly calculated by nonnegative least squares algorithm, $i.e.$,

$$\hat{\mathbf{X}}_\mathbf{s} = \arg\min \|\mathbf{Y} - \mathbf{A_s X_s}\|_F, \qquad s.t. \qquad \mathbf{X_s} \geq 0. \tag{11}$$

This constrained minimization problem can be simply solved by the Matlab function `lsqnonneg`. Put the obtained $\mathbf{X_s}$ into Eq. (10), the $\mathbf{A_s}$ could be further updated. Continue iterating until meet the stopping criterion. It is worth noting that k is usually unknown in practical, so we give it an estimation by HySime (Bioucas-Dias and Nascimento 2008) in this paper. HySime is just an available choice, and other similar algorithms such as Virtual Dimensionality (Chang and Du 2004) could also be used.
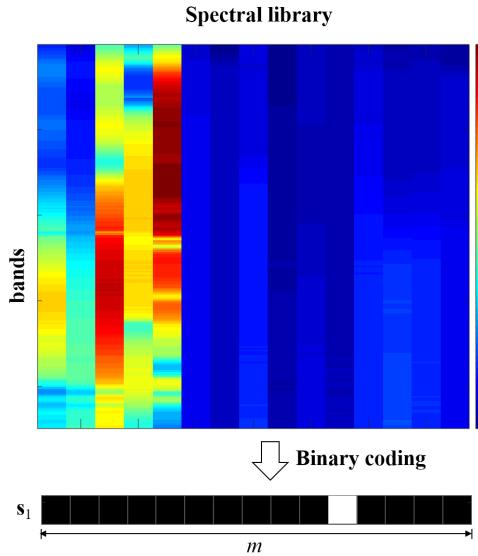
Fig. 1: Binary coding of spectral library. Here we simply suppose the spectral library contains m=15 spectra. The binary vector $\mathbf{s} = [s_1, ..., s_m]$ is the individuals. White points denote that this endmember is selected, and vice versa in black ones.

In this paper, we propose a new sparse unmixing method based on the MOEA/D framework. Firstly, we directly conduct MOEA/D for sparse unmixing (termed as MOEA/D-SU) in Section 3.1. Under this circumstance, a set of non-dominated solutions could be obtained to balance the

10

reconstruction error and sparsity error. However, we must determine a unique solution for P_{mo}. Thus in Section 3.2, we propose a new algorithm SMoSU by integrating the spectral characteristic of hyperspectral image into P_{te}. SMoSU takes the spectral correlation information among different individuals as a regularizer of the decomposition function. So the diversity and convergence of the population could be well balanced, which is beneficial to the concentration of individuals. In SMoSU, neither priori nor posterior operations about the solution set is required. The decision making problem is solved in the evolution process.

3.1. MOEA/D-SU

MOEA/D transforms the bi-objective optimization based sparse unmixing problem (P_{mo}) to a number of scalar single-objective optimization subproblems by using aggregate function. Then the subproblems are solved simultaneously to update the population in the evolution process iteration by iteration. In detail, a weight vector is assigned to each individual at the beginning, which is used to control the trajectory and searching direction of individuals. During evolution, the population is updated by the optimal solutions of current subproblems, where each subproblem is closely correlative to its neighboring subproblems. The neighbors of each subproblem are composed of individuals whose Euclidean distances of aggregate coefficient are similar. Tchebycheff approach is used here, which is based on Tchebycheff formula and is one of the most commonly used decomposition methods. Suppose the population size is p. The ith subproblem of (P_{mo}) is

$$
(P_{te}) : \min_{\mathbf{s}_i} g_i^{te}(\mathbf{s}_i | \boldsymbol{\lambda}_i, \mathbf{z}^*), \quad s.t. \ 0 < \|\mathbf{s}_i\|_1 < 2k
$$

$$
\text{where} \quad g_i^{te}(\mathbf{s}_i | \boldsymbol{\lambda}_i, \mathbf{z}^*) = \max_{1 \leq j \leq 2} \{\lambda_i^j | f_j(\mathbf{s}_i) - z_j^*|\}
$$

(12)

$\mathbf{z}^* = [z_1^*, z_2^*]^{\mathrm{T}}$ is the ideal point in current iteration with $z_j^* = \min\{f_j(\mathbf{s}_1), ..., f_j(\mathbf{s}_p)\}$, $\boldsymbol{\lambda}_i = [\lambda_i^1, \lambda_i^2]^{\mathrm{T}}$ is the weight vector with $\boldsymbol{\lambda}_i \geq 0$ and $\boldsymbol{\lambda}_i^{\mathrm{T}} \mathbf{1} = 1$, λ_i^j is the weight of the jth objective, $g_i^{te}(\mathbf{s}_i | \boldsymbol{\lambda}_i, \mathbf{z}^*)$ is the weighted Tchebycheff distance of individual $\mathbf{s}_i$ to the ideal point $\mathbf{z}^*$. The neighbors of the ith subproblem are subproblems whose weight vector is close to $\boldsymbol{\lambda}_i$. This approach could solve problems with any Pareto front shapes. The gap among individuals and the ideal point is reduced by (P_{te}) to achieve the convergence of (P_{mo}). And the optimal solution of the scalar problem (P_{te}) is one Pareto optimal solution of (P_{mo}). Although the non-Pareto based MOEA/D could bring high-efficiency to (P_{mo}), the large search space and computation load brought by over-complete spectral library are still serious. In fact, MOEAs are likely to trap into local optimum on high-

11

dimensional problems. Thus in this paper, the spectral library is pruned to a relatively small scale using RMUSIC (Fu et al. 2015).

Based on Eq. (10) and Eq (12) the MOEA/D-SU method could be constructed. Please find more detailed description about MOEA/D from (Zhang and Li 2007).

3.2. SMoSU

In the above original MOEA/D framework based sparse unmixing algorithm MOEA/D-SU, a set of non-dominated solutions could be obtained until the stopping criteria is satisfied. However, the decision of final solution is a serious problem of MOEA/D-SU. Therefore, in this paper, we consider integrating the spectral characteristics of hyperspectral images into the MOEA/D framework to solve the problem well and try to make the SMoSU concentrates to a single solution to avoid the decision making problem. We propose an improved Tchebycheff decomposition approach (P_{st}) by introducing an spectral information divergence (SID) based regularization term to (P_{te}). Then the ith subproblem of (P_{mo}) is

$$
\begin{aligned}
(P_{st}) : &\min_{\mathbf{s}_i} g_i^{st}(\mathbf{s}_i|\boldsymbol{\lambda}_i, \mathbf{z}^*, \mathbf{s}^*), \quad s.t.\ 0 < \|\mathbf{s}_i\|_1 < 2k \\
&\text{where}\quad g_i^{st}(\mathbf{s}_i|\boldsymbol{\lambda}_i, \mathbf{z}^*, \mathbf{s}^*) = \max_{1\leq j\leq 2}\{\lambda_i^j|f_j(\mathbf{s}_i) - z_j^*|\} + \mu \cdot \mathrm{SID}(\mathbf{A}_{\mathbf{s}_i}, \mathbf{A}_{\mathbf{s}^*})
\end{aligned}
\tag{13}
$$

$\mathbf{z}^*$ is the ideal point that is closest to the origin, $\mathbf{s}^*$ is the binary vector of $\mathbf{z}^*$, $\mathbf{A}_{\mathbf{s}_i}$ and $\mathbf{A}_{\mathbf{s}^*}$ are the corresponding spectra of individual $\mathbf{s}_i$ and $\mathbf{s}^*$, $\mathrm{SID}(\mathbf{A}_{\mathbf{s}_i}, \mathbf{A}_{\mathbf{s}^*})$ is the spectral information divergence of $\mathbf{A}_{\mathbf{s}_i}$ and $\mathbf{A}_{\mathbf{s}^*}$, μ is the regularization parameter. The whole process of SMoSU is shown in **Algorithm 1**.

3.2.1. SID of $\mathbf{A_s}$ and $\mathbf{A_{s^*}}$

Let $\mathbf{s} = [s_1, ..., s_{|\mathbf{s}|}]$, $\mathbf{s}^* = [s_1^*, ..., s_{|\mathbf{s}^*|}^*]$ and $s_1 \in \mathbf{s}$, $s_2 \in \mathbf{s}^*$, the relative entropy of $\mathbf{A}_{s_1}$ and $\mathbf{A}_{s_2}$ are

$$
\begin{aligned}
D(\mathbf{A}_{s_1}\|\mathbf{A}_{s_2}) &= \sum_{i=1}^{L} p_i\log(\frac{p_i}{q_i}) \\
D(\mathbf{A}_{s_2}\|\mathbf{A}_{s_1}) &= \sum_{i=1}^{L} q_i\log(\frac{q_i}{p_i})
\end{aligned}
\tag{14}
$$

where

$$
p_i = \frac{\mathbf{A}_{s_1}^i}{\sum_{i=1}^{L}\mathbf{A}_{s_1}^i}, q_i = \frac{\mathbf{A}_{s_2}^i}{\sum_{i=1}^{L}\mathbf{A}_{s_2}^i}
\tag{15}
$$

12

Algorithm 1: SMoSU

Input: hyperspectral image data $\mathbf{Y}$, spectral library $\mathbf{A}$.

Output: abundance fractions $\mathbf{X}$.

1 **Preprocessing**:

2 Estimate the number of active endmembers as $\hat{k}$ of the hyperspectral image using HySime;

3 Prune the spectral library $\mathbf{A}$ to a size m using RMUSIC.

4 **Initialization**:

5 population size p, neighborhood size ns, maximum iteration number T, regularization
   parameter μ, a population $\mathbf{S} = \{\mathbf{s}_1, ..., \mathbf{s}_p\}$, a set of weight vector $\mathbf{\Lambda} = \{\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_p\}$, indexes of
   each subproblem's neighbors $\{B_1, ..., B_p\}$, the ideal point $\mathbf{z}^* = [z_1^*, z_2^*]^{\mathrm{T}}$, an population
   archive EP.

6 **Endmember Selection**:

7 **while** $t < T$ **do**

8 $\quad$ $t = t + 1$;

9 $\quad$ **if** $t > 0.9 * T$ **then**

10 $\quad\quad$ Set $\mu = 0$

11 $\quad$ **for** $i = 1, ..., p$ **do**

12 $\quad\quad$ Generate a new individual $\mathbf{s}_i'$ from $\mathbf{s}_i$ based on bit-wise flipping strategy, where each
      bit is flipped with a probability $1/m$;

13 $\quad\quad$ **if** $\|F(\mathbf{s}^*)\|_2 > \|F(\mathbf{s}_j')\|_2$ **then**

14 $\quad\quad\quad$ Set $z^* = F(\mathbf{s}_i')$

15 $\quad\quad$ **for** $j \in B_i$ **do**

16 $\quad\quad\quad$ **if** $g_i^{st}(\mathbf{s}_i' | \boldsymbol{\lambda}_i, \mathbf{z}^*, \mathbf{s}^*) \leq g_j^{st}(\mathbf{s}_j | \boldsymbol{\lambda}_j, \mathbf{z}^*, \mathbf{s}^*)$ **then**

17 $\quad\quad\quad\quad$ Set $\mathbf{s}_j = \mathbf{s}_i'$ and $F(\mathbf{s}_j) = F(\mathbf{s}_i')$

18 $\quad$ Update the population archive EP.

19 Return the solution in EP as $\mathbf{s}^\star$ and record the corresponding spectral signatures.

20 **Abundance Estimation**:

21 Compute the abundances for the whole hyperspectral image based on nonnegative least
   squares algorithm: $\mathbf{X} = \arg\min_{\mathbf{X} \geq 0} \|\mathbf{Y} - \mathbf{A}_{\mathbf{s}^\star} \mathbf{X}\|_{\mathrm{F}}$

Then the SID of $\mathbf{A}_{s_i}$ and $\mathbf{A}_{s_i^*}$ is

$$\mathrm{sid}(\mathbf{A}_{s_1}, \mathbf{A}_{s_2}) = D(\mathbf{A}_{s_1} \| \mathbf{A}_{s_2}) + D(\mathbf{A}_{s_2} \| \mathbf{A}_{s_1}) \tag{16}$$

Based on $\mathrm{sid}(\mathbf{A}_{s_1}, \mathbf{A}_{s_2})$, a matrix $\widetilde{\mathrm{SID}}(\mathbf{A_s}, \mathbf{A_{s^*}})$ for all spectra of $\mathbf{A_s}$ and $\mathbf{A_{s^*}}$ can be obtained as

$$\widetilde{\mathrm{SID}}(\mathbf{A_s}, \mathbf{A_{s^*}}) =$$

$$\begin{bmatrix} \mathrm{sid}(\mathbf{A}_{s_1}, \mathbf{A}_{s_1^*}) & \cdots & \mathrm{sid}(\mathbf{A}_{s_1}, \mathbf{A}_{s_{\|\mathbf{s^*}\|_1}^*}) \\ \vdots & \ddots & \vdots \\ \mathrm{sid}(\mathbf{A}_{s_{\|\mathbf{s}\|_1}}, \mathbf{A}_{s_1^*}) & \cdots & \mathrm{sid}(\mathbf{A}_{s_{\|\mathbf{s}\|_1}}, \mathbf{A}_{s_{\|\mathbf{s^*}\|_1}^*}) \end{bmatrix}_{\|\mathbf{s}\|_1 \times \|\mathbf{s^*}\|_1} \tag{17}$$

In this paper, we define the SID of $\mathbf{A_s}$ and $\mathbf{A_{s^*}}$ as

$$\mathrm{SID}(\mathbf{A_s}, \mathbf{A_{s^*}}) = \frac{1}{2} \left(\frac{\sum_{i=1}^{\|\mathbf{s}\|_1} \|\widetilde{\mathrm{SID}}_{i,:}\|_\infty}{\|\mathbf{s}\|_1} + \frac{\sum_{i=1}^{\|\mathbf{s^*}\|_1} \|\widetilde{\mathrm{SID}}_{:,i}\|_\infty}{\|\mathbf{s^*}\|_1} \right) \tag{18}$$

where $\| \cdot \|_\infty$ is the infinite norm of a vector. It is used to measure the information divergence of spectra corresponding to two individuals.

3.2.2. Ideal point

In the original MOEA/D, the ideal point is virtual (we mark it as $\mathbf{z}_M^*$ here to distinguish it from the ideal point of SMoSU), defined by the following equation:

$$\mathbf{z}_M^* = [z_1^*, z_2^*]^{\mathrm{T}}, z_i^* = \min\{f_i(\mathbf{s}_j), i = \{1, 2\}, j = \{1, ..., p\}\} \tag{19}$$

$\mathbf{z}_M^*$ is located in the position where both f_1 and f_2 achieve minimums simultaneously, and all the individuals tend to get close to this virtual point.

However, in the problem of hyperspectral unmixing, the solution corresponds to a series of spectra. An immediate idea is that besides the weighted Tchebycheff distance, the spectral similarity between individuals and ideal point should also be considered. Individuals whose spectra are less similar to the ideal point are more likely to be replaced. If we still use virtual point as the ideal one, it may not even exist spectra, and this situation is certainly true if the Pareto front has come close to the edge of the range. If the ideal point does not contain real spectra, it is impossible and meaningless to compare the SID. Therefore, in SMoSU, the ideal point $\mathbf{z}^*$ is selected from the solution set in current iteration, which is defined by:

$$\mathbf{z}^* = F(\mathbf{s}_{i^*}), \quad s.t. \quad i^* = \arg\min_i \{\|F(\mathbf{s}_i)\|, i = 1, ..., p\} \tag{20}$$

14

In this case, the obtained $\mathbf{z}^*$ is a real point in the range which corresponds to real spectra.

In Fig. 2, we display an illustration about the difference between SMoSU and the original MOEA/D in the evolution process for ideal point determination. For the convenience of analysis, the range is illustrated as a continuous region. In Fig. 2, the blue region is a part of the range. All the solid points (eight yellow points and one red points) represents the population distribution in current iteration. In MOEA/D, the ideal point is determined as a virtual point, denoted by $\mathbf{z}^*_M$ in Fig. 2. In SMoSU, the ideal point is determined as an specific individual in the current population, denoted by $\mathbf{z}^*$ in Fig. 2.
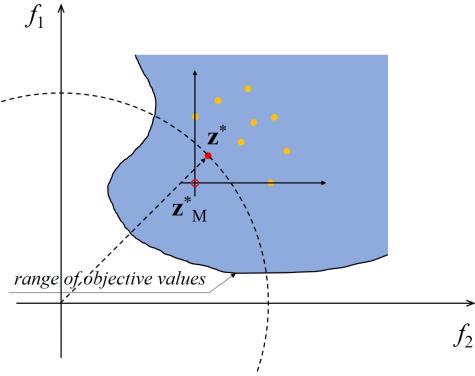
Fig. 2: An illustration about the difference of ideal points in the original MOEA/D (problem (P_{te})) and SMoSU ((P_{st})). $\mathbf{z}^*_M$ is the ideal point in (P_{te}). $\mathbf{z}^*$ is the ideal point in (P_{st}). The blue region is a part of the range. All the solid points (eight yellow points and one red points) represents the population distribution in current iteration. z^*_M and z^* are the ideal points in MOEA/D and SMoSU respectively.

3.2.3. Concentration of individuals

After sufficient iterations, the final optimization results are shown in Fig. 3. Each rectangular-box corresponds to a certain weighted Tchebycheff problem. Each point (solution) in curve $\overset{\frown}{AB}$ corresponds to a particular weight vector ($\boldsymbol{\lambda}$), and has minimum weighted Tchebycheff distance to the virtual ideal point. Curve $\overset{\frown}{AB}$ is just the Pareto optimal front in MOEA/D, which means all the points in $\overset{\frown}{AB}$ are Pareto optimal. By comparison, in SMoSU the ideal point actually exists in the range, denoted by "C". OC is perpendicular to the tangent of "C". Because "C" is located in the Pareto front, its effectiveness could be guaranteed. In SMoSU, for all individuals in the population, the minimum distance g^{st} to the ideal point is itself. Thus more and more individuals are replaced

by the ideal point, and the whole population are concentrated to a single solution gradually. In this way the decision making problem in hyperspectral unmixing is solved to some extent.
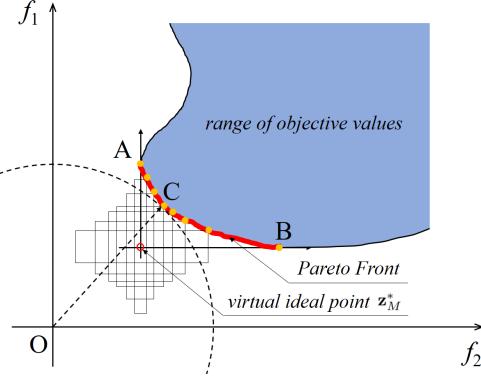
Fig. 3: An illustration about the difference between SMoSU and the original MOEA/D in the final optimization for ideal point determination. The blue region is a part of the range. Yellow points are some of the optimal solutions. $\widehat{\text{AB}}$ is the Pareto optimal front for MOEA/D. "C" is the ideal point in SMoSU.

3.2.4. Bit-wise flipping strategy

Bit-wise flipping is a mutation method, which is used to generate new individuals. Multi-objective optimization is a multiple variable based method, so a set of binary variables $\mathbf{S} = \{\mathbf{s}_1, ..., \mathbf{s}_p\}$ are initialized at the beginning (as shown in **Algorithm 1**). In each iteration, p new individuals $\{\mathbf{s}'_1, ..., \mathbf{s}'_p\}$ are generated. Each $\mathbf{s}'_i (i = 1, ..., p)$ is produced by the individual $\mathbf{s}_i$ based on the bit-wise flipping strategy. This strategy flips each bit (location) in $\mathbf{s}_i$ to its opposition with a probability $1/m$.

4. Experiments

In this section, synthetic and real-world experiments are designed to validate the performance of SMoSU respectively. The spectral library is Chapter 1 of the United States Geological Survey (USGS) digital spectral library (splib06a)[1], which contains 498 spectra under 224 bands distributed evenly in $0.4 - 2.5 \mu m$. Firstly, the performance analysis of SMoSU is presented, including the effect of the regularization parameter μ, SRE results under different k with $\hat{k}$ ranging from 3

---

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

to 10, the concentration process of SMoSU. Secondly, SMoSU is compared with five state-of-art algorithms: SUnSAL (Bioucas-Dias and Figueiredo 2010), SUnSAL-TV (Iordache et al. 2012), SMP (Shi et al. 2014), RSFoBa-2, RSFoBa-Inf (Tang et al. 2014) on three synthetic data sets and one real-world data. The signal-to-reconstruction error ($\mathrm{SRE} \equiv 10\log_{10}\left(\mathrm{E}[\|\mathbf{X}\|_F^2]/\mathrm{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2]\right)$) is used to evaluate the performance of algorithms, where $\mathbf{X}$ is the true abundances and $\hat{\mathbf{X}}$ is the estimated ones. Generally, larger SRE values correspond to better accuracy. Because SMoSU actually works as a subset selection method, as long as the exact endmembers are found, their SREs should be the same. For parameters setting, the default values in MOEA/D are taken for population size p and neighborhood size n_s, because SMoSU is not sensitive to these parameters. The maximum iteration number T is set as 100 both in synthetic and real-world experiments based on the results in Table 2, and the library size m after prunning is 40 according to (Iordache et al. 2014b).

Readers can directly use the above default parameters to handle other preferred test data, and the Matlab code of SMoSU is published online[2].

4.1. Performance analysis of SMoSU

4.1.1. Effect of the regularization parameter μ

It has been known that the ℓ_q $(0 < q \leq 1)$ norm relaxed sparse unmixing algorithms are sensitive to regularization parameters in (P_1). Thus it is necessary to analyse the effect of μ in (P_{st}). We test μ on a synthetic data generated by 10 spectra selected from the USGS library. The abundance fractions are generated based on Dirichlet distribution and forced to be smaller that 0.7 to avoid pure pixels. Because the real noise form is unavailable, here we add relatively complex correlated noise for better simulation with 20/30/40dB signal-to-noise ratios ($\mathrm{SNR} \equiv 10\log_{10}\left(\|\mathbf{Y}\|_F^2/\|\mathbf{N}\|_F^2\right)$). The correlated noise is generated by conducting low pass filtering on i.i.d. Gaussian noise with a normalized cutoff frequency of $5\pi/L$ (Iordache et al. 2011). Table 2 presents the SRE results, processing-times and maximum iterations T of SMoSU with μ ranging from 0 to 1. If the maximum iteration number is larger than 100, the results are not taken into account. In fact, SMoSU usually can still converge in these situations. The values of SRE present a little fluctuation when $T < 100$. It can be observed from Table 2 that the added term mainly works in the high-noise conditions.

---

[2]http://levir.buaa.edu.cn/Code.htm

17

When the noise is not strong (30/40dB), it is relatively easy for SMoSU to find the optimal solution. So it is not necessary to further integrating the SID term. However, the time cost varies in different μ. SMoSU can obtain the best results when $\mu = 0.2$ in the 20dB noise cases, and $\mu = 0.6 / \mu = 0.8$ for 30/40dB noise cases respectively. Therefore, considering greater effect of μ in 20dB cases, μ is set to 0.2 in all the following experiments.

Table 2: SRE results, processing-times, and maximum iterations T of SMoSU with μ ranging from 0 to 1 under 20/30/40dB noise level. The best results are in bold.

SNR	μ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	SRE	4.63	5.56	**5.56**	3.63	5.60	4.63	4.63	4.63	4.63	4.63	4.63
20dB	Time(s)	366.7	382.2	**357.3**	456.6	458.3	493.8	473.1	502.3	545.6	627.3	936.7
	T	73	76	**69**	90	89	110	107	111	123	138	217
	SRE	18.35	18.35	18.35	18.35	18.35	18.35	**18.35**	18.35	18.35	18.35	18.35
30dB	Time(s)	245.1	221.9	212.0	181.9	185.6	194.6	**180.9**	211.9	220.3	256.3	364.8
	T	66	61	59	49	49	49	**48**	58	60	80	114
	SRE	27.46	27.46	27.46	27.46	27.46	27.46	27.46	27.46	**27.46**	27.46	27.46
40dB	Time(s)	169.2	230.5	232.6	199.9	164.5	162.3	166.5	162.1	**157.2**	166.4	207.4
	T	45	61	65	50	40	40	40	40	**40**	40	51

4.1.2. Effect of $\hat{k}$

According to Eq. (10), the performance of SMoSU is highly correlated to the estimated endmember number $\hat{k}$. Here, we test the effect of $\hat{k}$ on synthetic data sets corrupted by 30dB correlated noise. The number of actual endmembers k varies from 3 to 10 and $\hat{k}$ is also assumed to range from 3 to 10, which includes all $k < \hat{k}$, $k = \hat{k}$ and $k > \hat{k}$ cases. Fig. 4 illustrates the relation among k, $\hat{k}$ and the number of obtained endmembers by SMoSU. The values inside Fig. 4 are the differences between k and our obtained endmembers number. "0" denotes the obtained endmembers number is the same as k. However, only using Fig. 4 we cannot determine whether the obtained endmembers cover the actual ones. Thus Table 3 is presented to show the corresponding SRE for the cases in Fig. 4. In Table 3, SMoSU performs well when the endmember number is estimated accurately (i.e. $\hat{k} = k$) or a little smaller. Usually, SMoSU could select the exact endmembers in these cases.

In addition, the results fluctuate slightly when $\hat{k} > k$. We may conclude that obtained endmembers could cover the actual ones in these cases. However, when $\hat{k}$ is much smaller than k, the unmixing accuracy presents obvious decline. The results in Table 3 and Fig. 4 indicate that overestimation has little influence on the accuracy while missing active endmembers should be avoided.

Table 3: The SRE results of SMoSU under different $\hat{k}$ with the actual endmember number ranging from 3 to 10.

	$\hat{k}=3$	$\hat{k}=4$	$\hat{k}=5$	$\hat{k}=6$	$\hat{k}=7$	$\hat{k}=8$	$\hat{k}=9$	$\hat{k}=10$
$k=3$	**18.65**	18.37	18.35	18.35	17.37	18.38	16.86	16.87
$k=4$	19.42	**19.42**	18.62	18.09	17.20	17.74	15.80	14.92
$k=5$	4.20	17.65	**17.65**	17.34	16.54	14.52	14.80	14.88
$k=6$	5.24	18.01	18.01	**18.01**	17.91	16.39	17.43	15.92
$k=7$	2.73	17.58	17.58	17.58	**17.58**	17.29	17.08	13.28
$k=8$	1.33	6.29	17.60	17.60	17.60	**17.60**	13.51	12.65
$k=9$	-4.43	-3.81	6.77	18.03	18.03	18.03	**18.03**	14.26
$k=10$	-0.55	2.76	7.33	7.33	7.35	18.35	18.35	**18.35**

4.1.3. Concentration process of SMoSU

SMoSU avoids the decision making problem in MOEA/D-SU by integrating the spectral characteristic of hyperspectral images into the original MOEA/D framework. In order to verify the effectiveness of this improvement, Fig. 5 and Fig. 6 are presented. In Fig. 5, the objective function values of individuals after 20/40/60/80 iterations in MOEA/D-SU and SMoSU are compared respectively. These experiments are conducted on synthetic data generated by 10 spectra with 30dB correlated noise. The first row in Fig. 5 are solutions by MOEA/D-SU. We can see that after sufficient iterations the shape of Pareto front tends to stable, especially in the left side. However, there are many solutions in MOEA/D-SU (shown by blue circles) and thus the decision making problem exists. Results by SMoSU are shown in the second row. It is observed that the solutions tend to concentrate with the increase of iterations. Similar situation could also be observed in Fig. 6. The first column is our result and the second one corresponds to that of MOEA/D-SU. Take a sub-figure for example. The x-axis denotes all the 100 individuals, and the y-axis is serial number of the library. White points denote that this endmember is selected, and vice versa in black ones. Fig. 6 demonstrates that individuals in SMoSU have presented very consistent results in all the settings of k. However, individuals in MOEA/D-SU are diverse.

19
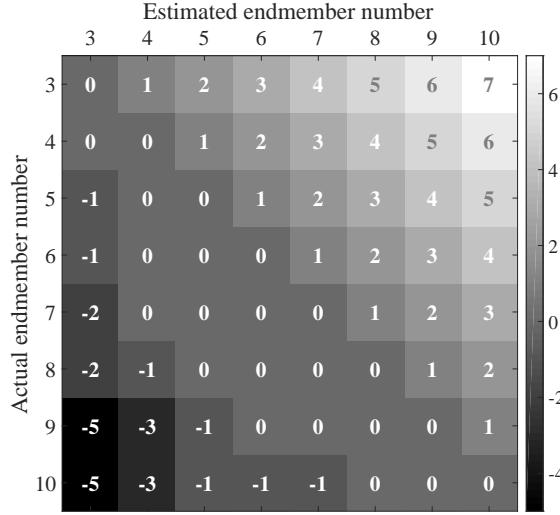
Fig. 4: Endmember missing and redundancy of SMoSU under different k and $\hat{k}$. Positive numbers indicate the redundant endmember numbers and negative number indicate the missing endmember numbers. The '0' cases indicate the number of selected endmembers are the same with the truth.
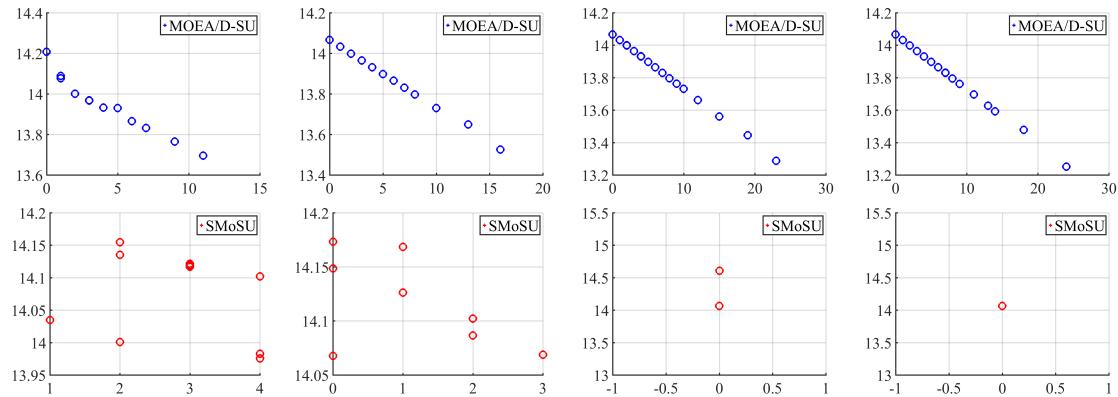
Fig. 5: The concentration process of MOEA/D-SU and SMoSU when the actual endmember number is 10 and noise is 30dB. x-axis and y-axis of subfigures denote the values of f_2 and f_1, respectively. The first row are the objective function values of MOEA/D-SU when the t=20/40/60/80. The second row are the objective function values of SMoSU when the t=20/40/60/80.

20
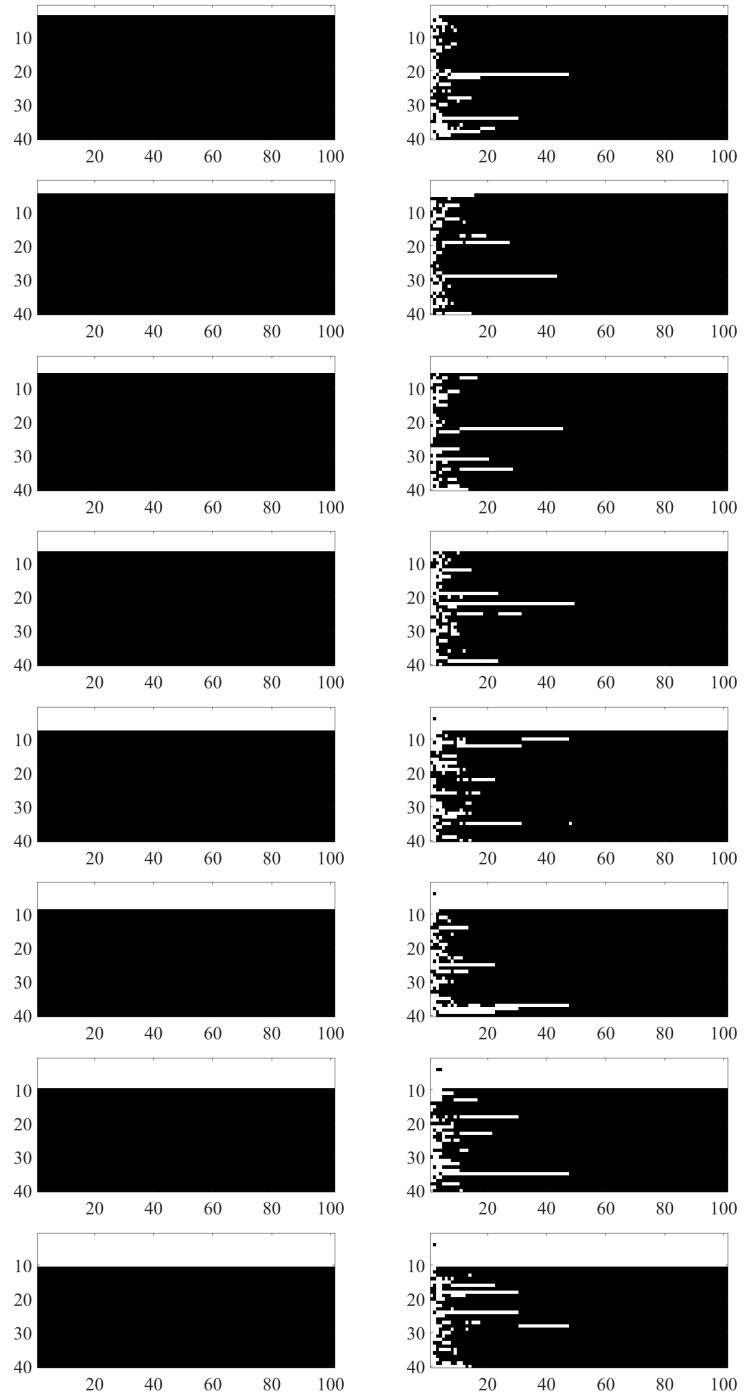
Fig. 6: The binary images of populations for SMoSU and MOEA/D-SU. The left column is populations of SMoSU, and the right column is populations of MOEA/D-SU. From top to bottom are the obtained populations for $k = 3$ to $k = 10$ respectively. For each subfigure, the x-axis denotes all the 100 individuals, and the y-axis is serial number of the library. White points denote that this endmember is selected, and vice versa in black ones.

4.2. Comparing with state-of-art algorithms

4.2.1. Synthetic data 1

Synthetic data 1 contains images of size 64×64, which are generated by 3 to 10 spectral signatures with 20/30/40dB correlated noise respectively. The abundances follows a Dirichlet distribution and are forced to be smaller than 0.7 to avoid pure pixels. To increase the unmixing difficulty, we specially choose 5 similar spectra from the library: Actinolite HS116.3B, Actinolite HS22.3B, Actinolite HS315.4B, Actinolite NMNH80714 and Actinolite NMNHR16485. Table 4 shows the results obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf and SMoSU with different k and different noise level. Generally speaking, the SRE results of all algorithms are decreased with the increase of noise strength. The results of SMoSU fluctuates slightly with k varying from 3 to 10. SUnSAL-TV performs worse than SUnSAL in most cases, since the Dirichlet distribution of abundances may mislead the spatial context constraint of SUnSAL-TV. In the 20dB noise cases, SMoSU surpasses the five state-of-art algorithms in most cases. It is slightly worse than the best SRE when the actual endmember number is 3 and 6, but outperforms the others when k is large. The best results are obtained by SMoSU in all the 30/40dB cases. In fact, all actual endmembers are exactly selected in theses situations. It is worth mentioning that each of the five state-of-art algorithms have several parameters to adjust according to different noise level and k, which leads to a great workload. However, for SMoSU, the parameters are usually robust and do not need to be adjusted.

4.2.2. Synthetic data 2

Synthetic data 2 is introduced by Iordache et al. (2012), which is generated by 5 spectral signatures with a 75×75 image size and has a good spatial homogeneity. It is also corrupted by 20/30/40dB correlated noise respectively. Table 5 and Fig. 7 give quantitative and visual comparison of the algorithms. In Table 5, the SRE results of the six algorithms on different noise levels are listed. Fig. 7 presents the abundance maps of these algorithms with 30dB correlated noise. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU and the truth respectively. From top row to bottom row are the maps corresponding to endmembers #1 to #5. It can be observed that SUnSAL-TV presents better spatial smoothness, but loses some small patches for endmembers #1 to #4. Table 5 is more convincing since it is not easy to justify which method performs better from visual results. Note that

22

Table 4: SRE results of SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf and SMoSU on synthetic data 1.

SNR	k	SUnSAL	SUnSAL-TV	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU
20dB	3	4.5052	5.2438	**8.6130**	8.4044	**8.6130**	7.7487
	4	3.2859	2.7137	2.5247	7.9220	7.4005	**9.3768**
	5	3.0177	2.3333	3.2845	8.5611	7.4656	**8.8596**
	6	3.3636	2.8785	5.7019	5.1444	**7.1982**	5.0969
	7	3.2529	2.9369	4.0000	3.8065	5.7600	**6.1454**
	8	2.3822	2.2278	3.9736	4.6277	3.9579	**5.4857**
	9	2.7601	2.6670	3.8935	3.8254	4.0621	**6.6004**
	10	3.1040	2.8880	4.7120	3.2561	3.6430	**5.5628**
30dB	3	11.3329	11.5359	18.5585	18.5585	**18.6567**	**18.6567**
	4	9.5467	8.6930	3.5312	16.8542	16.1260	**19.4240**
	5	9.0219	7.6637	4.9450	14.2502	15.0515	**17.6580**
	6	9.2247	7.9197	5.5319	13.8184	14.2658	**18.0181**
	7	8.8378	7.8053	6.2037	12.7791	14.2503	**17.5835**
	8	7.9069	7.1904	6.2300	11.5081	12.9609	**17.6047**
	9	8.1637	7.5223	7.2575	12.2967	13.0632	**18.0314**
	10	8.8004	8.0950	7.4321	11.4065	12.9061	**18.3527**
40dB	3	19.6442	21.9383	28.2405	28.2405	28.2405	**28.3241**
	4	16.0227	16.0821	3.7792	25.6794	25.6794	**29.3713**
	5	15.3377	15.0203	5.0044	23.0806	24.5285	**27.2786**
	6	15.2891	15.0866	6.1201	22.4171	24.2171	**27.3813**
	7	14.7076	14.8440	6.6270	24.0554	23.2697	**26.9648**
	8	14.0966	14.4336	7.2206	22.1368	21.2189	**27.0626**
	9	14.6300	14.8271	7.7812	19.9435	22.0977	**27.0128**
	10	14.7714	15.1587	8.3255	19.4626	20.5550	**27.4649**

although it looks like SMoSU has achieved much higher SREs than other, its real advantage is not very significant. These large gaps are mainly generated due to the characteristic when calculating SREs.

Table 5: SRE results of SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf and SMoSU on synthetic data 2.

SNR	SUnSAL	SUnSAL-TV	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU
20dB	2.8740	7.2579	5.2260	4.2532	5.5423	**9.2787**
30dB	6.1289	15.0983	13.5314	12.3136	13.4010	**15.8934**
40dB	11.0483	21.4753	23.9048	22.7590	23.9048	**26.4514**

4.2.3. Synthetic data 3

Synthetic data 3 is also a commonly used dataset (Iordache et al. 2012). It is generated by 9 spectra and contains 100×100 pixels. Table 6 and Fig. 8 present the results of the six algorithms, which are organized by similar manner as Table 5 and Fig. 7. The advantage of SMoSU mainly reflects on the 20dB data, according to Table 6. Actually, SMoSU could find the exact endmembers in 30dB and 40dB data, and thus it present closer results to the truth.

Table 6: SRE results of SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf and SMoSU on synthetic data 3.

SNR	SUnSAL	SUnSAL-TV	SMP	RSFoBa-2	RSFoBa-Inf	SMoSU
20dB	3.5823	6.8356	11.7575	9.1401	10.4579	**12.3412**
30dB	8.0323	11.8903	15.2202	16.5553	19.7419	**22.2353**
40dB	12.9896	17.1311	21.6863	25.0255	27.1758	**33.8674**
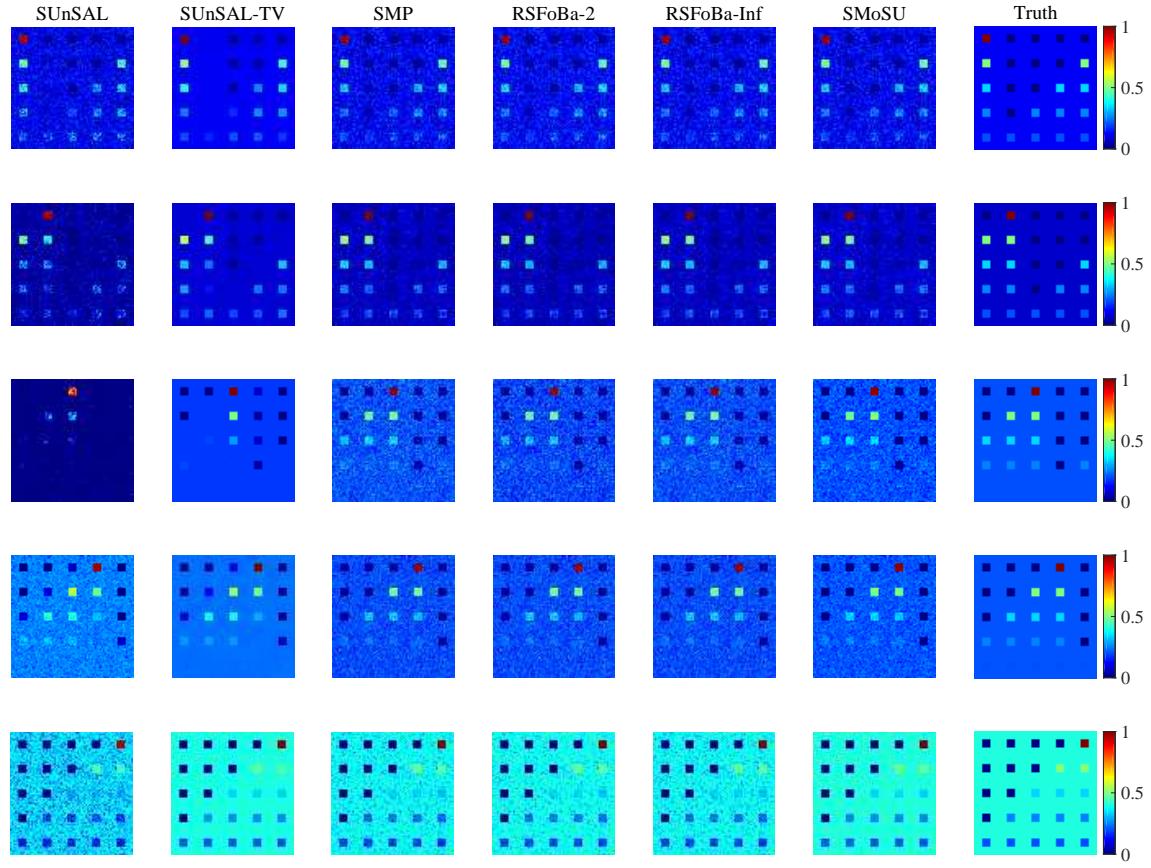
Fig. 7: Comparison of abundance maps on synthetic data 2 with 30dB correlated noise. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU and the truth respectively. From top row to bottom row are the maps corresponding to endmembers #1 to #5.
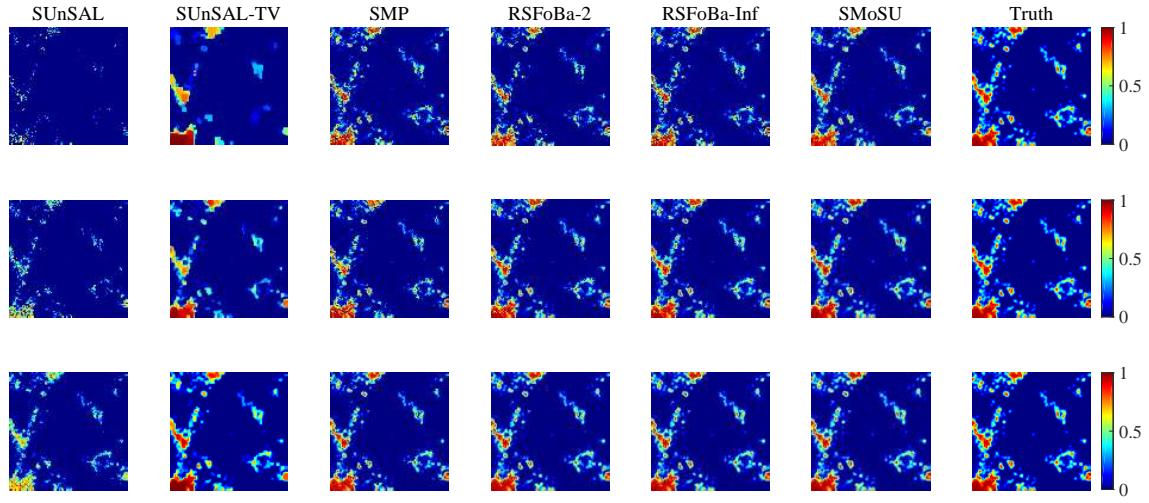
Fig. 8: Abundance maps of endmember #3 on synthetic data 3 with 20/30/40dB correlated noise. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf, SMoSU and the truth respectively. From top row to bottom row are the maps corresponding to 20/30/40dB correlated noise.
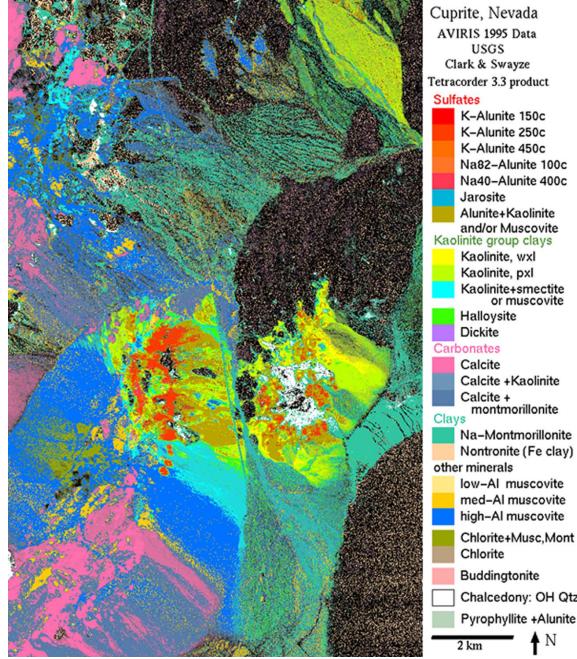
Fig. 9: Material map of Cuprite dataset obtained by Tricorder 3.3 software.
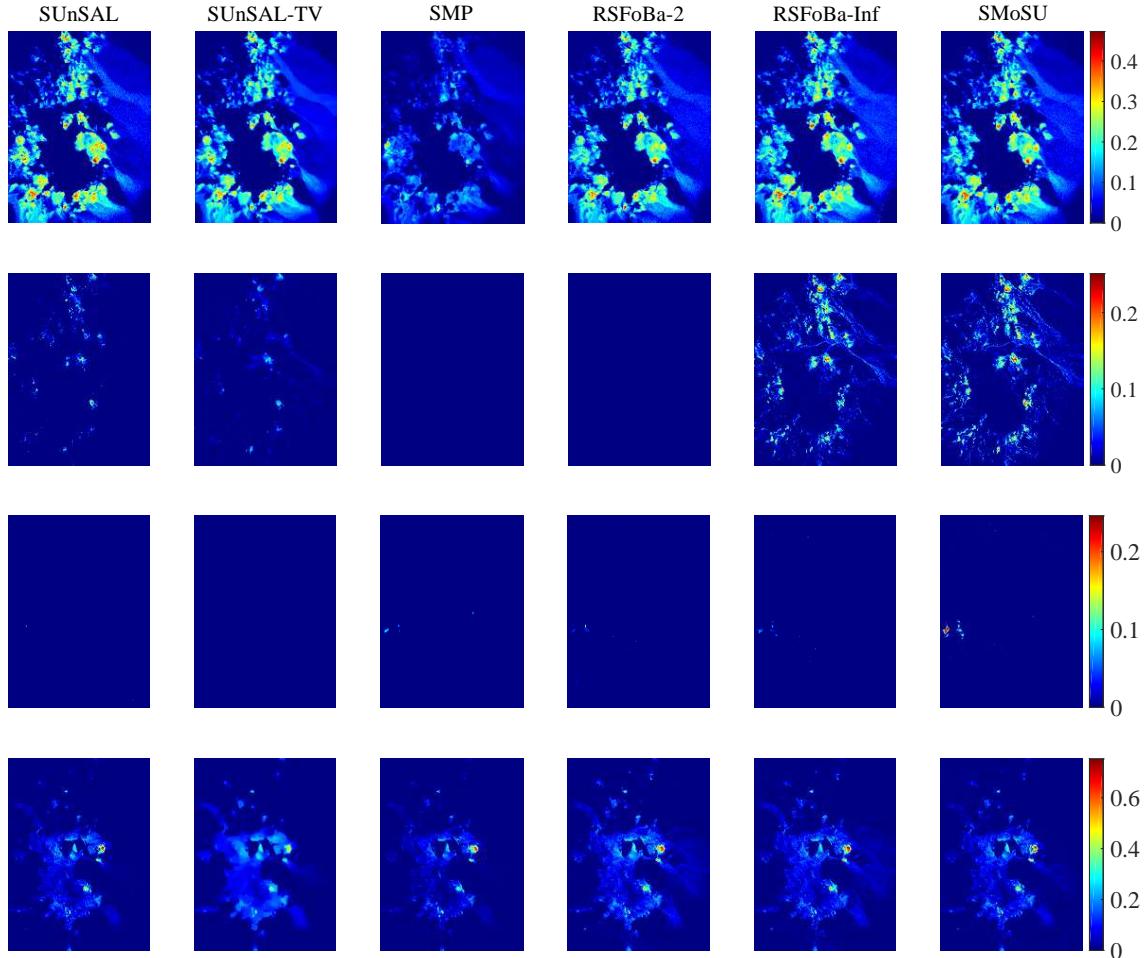
Fig. 10: Comparison of abundance maps on Cuprite data. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf and SMoSU respectively. From top row to bottom row are the maps corresponding to Alunite+Muscovite/Kaolinite, Alunite, Ammonio-Smectite and Chalcedony.

4.2.4. Real-world data

The real-world data we used here is a 204×151 subscene of the well-known AVIRIS Cuprite dataset[3]. This dataset contains 188 spectral bands after a removing of water absorption bands and some low SNR bands. A material map[4] produced by Tricorder 3.3 software is shown in Fig. 9, where each pixel is labeled pure and classified to a certain material. This is conflicting to the mixing premise of sparse unmixing. Thus Fig. 9 can only give a qualitative analysis for the unmixing results of all the six algorithms. We divide the subscene into several blocks in this experiment, since a large number of materials are contained in the real-world image. Fig. 10 presents a comparison of these algorithms on the Cuprite dataset. From left column to right column are abundance maps obtained by SUnSAL, SUnSAL-TV, SMP, RSFoBa-2, RSFoBa-Inf and SMoSU respectively. From top row to bottom row are the maps corresponding to Alunite+Muscovite/Kaolinite, Alunite, Ammonio-Smectite and Chalcedony. From Fig. 10, the abundance maps of SMoSU is similar with SUnSAL, SUnSAL-TV, RSFoBa-2 and RSFoBa-Inf for Alunite+Muscovite/Kaolinite. SMP and RSFoBa-2 fail to find Alunite and SMoSU surpasses all the five state-of-art algorithms for Ammonio-Smectite. In addition, all algorithms perform well for Chalcedony and obtain similar abundance maps. In conclusion, SMoSU is effective for real-world images.

5. Conclusion

In this paper, a multi-objective optimization based sparse unmixing (SMoSU) is proposed. SMoSU is developed based on MOEA/D framework. To solve the ℓ_0 norm sparse unmixing problem directly, spectral library is encoded in binary vectors. Each spectral signature is replaced by a binary code and thus sparse unmixing is transformed to a bi-objective discrete optimization problem. A bit-wise randomly flipping strategy is used to generate new individuals in the evolution process of SMoSU. In order to select the only optimal solution from the non-dominated solutions, we integrate spectral characteristic of hyperspectral image into SMoSU. A new regularizer which includes spectral correlation information is added to the Tchebycheff decomposition approach in SMoSU. In this way, the diversity and convergence of population is further balanced and the concentration of individuals are ensured. SMoSU do not need priori operations to specify a utility function over all different

---

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

objective functions nor posterior operations to make a decision from the solution set. It is worth mentioning that the proposed SMoSU method is a multi-objective based algorithm, rather than a complete multi-objective optimization one. MO problems try to obtain a solution set as diverse and uniform as possible, but SMoSU strives to obtain a single optimal solution. In general, our strategy can be directly used to improve other weighted-metrics based multi-objective methods.

To verify the performance of SMoSU, effectiveness analysis experiments and contrast experiments are conducted. We analyse the concentration process of SMoSU and test the effect of some key parameters, including the estimated endmember number and regularization parameter. Three synthetic and one real-world datasets are used to compare SMoSU with several state-of-art algorithms.

References

Jos M. Bioucas-Dias and Mario A. T. Figueiredo. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS 2010*, pages 1–4, 2010.

Jos M. Bioucas-Dias and Jos M. P. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008.

Jos M. Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

Chein-I Chang and Qian Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):608–619, March 2004.

Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014.

Kalyanmoy Deb and Deb Kalyanmoy. *Multi-objective optimization using evolutionary algorithms.* John Wiley and Sons, Inc, 2001.

Ruyi Feng, Yanfei Zhong, and Liangpei Zhang. Adaptive non-local euclidean medians sparse unmixing for hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 97: 9–24, 2014.

Ruyi Feng, Yanfei Zhong, and Liangpei Zhang. Adaptive spatial regularization sparse unmixing strategy based on joint MAP for hyperspectral remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12):5791–5805, 2016.

Xiao Fu, Wing Kin Ma, Jos M. Bioucas-Dias, and Tsung Han Chan. Semiblind hyperspectral unmixing in the presence of spectral library mismatches. *IEEE Transactions on Geoscience and Remote Sensing*, 54(9):5171–5184, 2015.

Maoguo Gong, Hao Li, Enhu Luo, Jing Liu, and Jia Liu. A multi-objective cooperative coevolutionary algorithm for hyperspectral sparse unmixing. *IEEE Transactions on Evolutionary Computation*, 21(2):234–248, 2017.

Pinghua Gong, Changshui Zhang, Zhaosong Lu, Jianhua Z. Huang, and Jieping Ye. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *30th International Conference on Machine Learning, ICML 2013*, pages 696–704, 2013.

Rob Heylen, Mario Parente, and Paul Gader. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 1844–1868, 2014.

Marian Daniel Iordache, Jos M. Bioucas-Dias, and Antonio Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2014–2039, 2011.

Marian Daniel Iordache, Jos M. Bioucas-Dias, and Antonio Plaza. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4484–4502, 2012.

Marian-Daniel Iordache, Jose M. Bioucas-Dias, and Antonio Plaza. Collaborative sparse regression for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):341–354, 2014a.

Marian Daniel Iordache, Jose M. Bioucas-Dias, Antonio Plaza, and Ben Somers. MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7):4364–4382, 2014b.

Himanshu Jain and Kalyanmoy Deb. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part II: Handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation*, 18(4):602–622, 2014.

Nirmal Keshava and John F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19 (1):44–57, 2002.

Lin Li, Xin Yao, Rustam Stolkin, Maoguo Gong, and Shan He. An evolutionary multiobjective approach to sparse reconstruction. *IEEE Transactions on Evolutionary Computation*, 18(6): 827–845, 2014.

A. Ma, Y. Zhong, and L. Zhang. Adaptive multiobjective memetic fuzzy clustering algorithm for remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4202–4217, Aug 2015.

A. Ma, Y. Zhong, D. He, and L. Zhang. Multiobjective subpixel land-cover mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 56(1):422–435, Jan 2018.

Wing Kin Ma, Jose M. Bioucas-Dias, Tsung Han Chan, and Nicolas Gillis. A signal processing perspective on hyperspectral unmixing: insights from remote sensing. *IEEE Signal Processing Magazine*, 31(1):67–81, 2014.

Kaisa Miettinen. *Nonlinear Multiobjective Optimization*, volume 12. Springer Science & Business Media, 1999.

Bin Pan, Zhenwei Shi, Ning Zhang, and Shaobiao Xie. Hyperspectral image classification based on nonlinear spectral–spatial network. *IEEE Geoscience and Remote Sensing Letters*, 13(12): 1782–1786, 2016.

Bin Pan, Zhenwei Shi, and Xia Xu. MugNet: Deep learning for hyperspectral image classification using limited samples. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017a. doi: https://doi.org/10.1016/j.isprsjprs.2017.11.003.

Bin Pan, Zhenwei Shi, and Xia Xu. R-VCANet: A new deep-learning-based hyperspectral image classification method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):1975–1986, 2017b.

Chao Qian, Yang Yu, and Zhi-Hua Zhou. On constrained boolean pareto optimization. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 389–395, 2015a.

Chao Qian, Yang Yu, and Zhi-Hua Zhou. Subset selection by pareto optimization. In *Advances in Neural Information Processing Systems*, pages 1774–1782, 2015b.

Chao Qian, Jing Cheng Shi, Yang Yu, Ke Tang, and Zhi Hua Zhou. Parallel pareto optimization for subset selection. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 1939–945, 2016.

Yuntao Qian, Sen Jia, Jun Zhou, and Antonio Robles-Kelly. Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4282–4297, 2011.

Frank J. A. Van Ruitenbeek, Pravesh Debba, Freek D. Van Der Meer, Thomas Cudahy, Mark Van Der Meijde, and Martin Hale. Mapping white micas and their absorption wavelengths using hyperspectral band ratios. *Remote Sensing of Environment*, 102(3–4):211–222, 2006.

Chen Shi and Le Wang. Incorporating spatial information in spectral unmixing: A review. *Remote Sensing of Environment*, 149:70–87, 2014.

Zhenwei Shi, Wei Tang, Zhana Duren, and Zhiguo Jiang. Subspace matching pursuit for sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6): 3256–3274, 2014.

Yan Sun, Derrick Wing Kwan Ng, Jun Zhu, and Robert Schober. Multi-objective optimization for robust power efficient and secure full-duplex wireless communication systems. *IEEE Transactions on Wireless Communications*, 15(8):5511–5526, 2016.

Wei Tang, Zhenwei Shi, and Ying Wu. Regularized simultaneous forward–backward greedy algorithm for sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5271–5288, 2014.

Qi Wang, Jianzhe Lin, and Yuan Yuan. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Transactions on Neural Networks and Learning Systems*, 27(6):1279–1289, 2016.

X. Wang, Y. Zhong, L. Zhang, and Y. Xu. Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6287–6304, Nov 2017.

Rebecca M. Willett, Marco F. Duarte, Mark A. Davenport, and Richard G. Baraniuk. Sparsity and structure in hyperspectral imaging: sensing, reconstruction, and target detection. *IEEE Signal Processing Magazine*, 31(1):116–126, 2014.

Xia Xu and Zhenwei Shi. Multi-objective based spectral unmixing for hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124:54–69, 2017.

Bing Xue, Mengjie Zhang, Will N. Browne, and Xin Yao. A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4):606–626, 2016.

Yang Yu, Xin Yao, and Zhi Hua Zhou. On the approximation ability of evolutionary optimization with application to minimum set cover: extended abstract. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 3190–3194, 2013.

Alina Zare and K.C. Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2014.

Qingfu Zhang and Hui Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

Ridong Zhang and Jili Tao. Data-driven modeling using improved multi-objective optimization based neural network for coke furnace system. *IEEE Transactions on Industrial Electronics*, 64 (4):3147–3155, April 2017.

Yanfei Zhong, Xinyu Wang, Lin Zhao, Ruyi Feng, Liangpei Zhang, and Yanyan Xu. Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119:49–63, 2016.

Yanfei Zhong, Ailong Ma, Yew Soon Ong, Zexuan Zhu, and Liangpei Zhang. Computational intelligence in optical remote sensing image processing. *Applied Soft Computing*, 64:75–93, 2018.

Aimin Zhou, Bo Yang Qu, Hui Li, Shi Zheng Zhao, Ponnuthurai Nagaratnam Suganthan, and Qingfu Zhang. Multi-objective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, 1(1):32–49, 2011.

Yicong Zhou and Yantao Wei. Learning hierarchical spectral–spatial features for hyperspectral image classification. *IEEE transactions on cybernetics*, 46(7):1667–1678, 2016.

Feiyun Zhu, Ying Wang, Bin Fan, Shiming Xiang, Geofeng Meng, and Chunhong Pan. Spectral unmixing via data-guided sparsity. *IEEE Transactions on Image Processing*, 23(12):5412–5427, 2014a.

Feiyun Zhu, Ying Wang, Shiming Xiang, Bin Fan, and Chunhong Pan. Structured sparse method for hyperspectral unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88:101–118, 2014b.