Simultaneously Multiobjective Sparse Unmixing and Library Pruning for Hyperspectral Imagery

Xia Xu, *Graduate Student Member, IEEE*, Bin Pan, *Member, IEEE*, Zongqing Chen, Zhenwei Shi, *Member, IEEE*, and Tao Li

***Abstract*— Sparse hyperspectral unmixing has attracted increasing investigations during the past decade. Recent research has indicated that library pruning algorithms can significantly improve the unmixing accuracies by reducing the mutual coherence of the spectral library. Inspired by the good performance of library pruning, in this article we propose a new hyperspectral unmixing algorithm which integrates the idea of library pruning and sparse representation. An obvious challenge for pruning algorithms is that the real endmembers must be preserved after pruning. Unfortunately, recent proposed pruning algorithms, such as multiple signal classification are actually prepruning strategies, which cannot guarantee that the endmembers exactly exist in the selected spectral subset when the image noise is strong. To overcome this difficulty, we develop a simultaneous optimization approach which involves the pruning operation into the optimization process. Compared with existing prepruning-based unmixing methods, the proposed algorithm can gradually compress the search space of sparse representation, which may relieve the loss of spectral information caused by the rapid compression of the library. Instead of simply designing a regularizer, in this article we utilize a multiobjective-based framework where reconstruction error, sparsity error, and the pruning projection function are considered as three parallel objectives, so as to avoid the manually settings of regularization parameters. Moreover, we have provided theoretical analysis and proof for the reasonability of our pruning objective. Experiments on synthetic hyperspectral data may indicate the superiority of the proposed method under high-noise conditions.**

***Index Terms*— Library pruning, multiobjective optimization, sparse hyperspectral unmixing.**

I. Introduction

HYPERSPECTRAL remote sensing imagery includes abundant spectral information which is quite beneficial to land-cover identification task. However, current imaging technology cannot improve spatial and spectral resolutions at the same time [1], thus hyperspectral imagery has to suffer relatively low spatial resolution. In this case, one pixel in hyperspectral imagery usually contains several materials, and this pixel is called "mixed" [2]. Spectral unmixing is the process that extracts the pure materials spectra (called *endmembers*) in the mixed pixels and estimates their fractional proportions (called *abundances*) [2], [3]. Correspondingly, endmember extraction and abundance inversion are two general cascade tasks during unmixing. According to different mixing assumptions, unmixing methods can be divided into linear and nonlinear mixing models. In this article, we adopt linear mixing assumption.

Recent linear unmixing methods can be roughly divided into four categories [3]: geometry, statistics, nonnegative matrix factorization, and sparse regression-based approaches. Geometrical-based methods usually try to maximize or minimize the simplex volume whose vertices are considered as the endmembers [4]–[6]. Statistical methods perform well when the materials are highly mixed [7]–[10], but they usually accompany with higher computational complexity. Nonnegative matrix factorization can divide the hyperspectral data into two nonnegative matrices [11]. With reasonable constraints and initialization, these two matrices can be transformed to endmember and abundance matrices [11]–[14].

Sparse regression-based unmixing (also called sparse unmixing) is a semisupervised fashion, where it assumes that the spectral signatures can be expressed in the form of linear combinations of a number of pure spectral signatures known in advance [15]. The prior spectral signature set is called spectral library. Sparse unmixing is immune to pure pixel constraint, and the computation cost is acceptable. Due to the above advantages, sparse unmixing has become a hot topic. Therefore, in this article our method is developed under the framework of sparse unmixing.

Sparse unmixing methods usually focus on three research aspects: sparsity handling, spatial information, and spectral variability. Sparse unmixing is a combinational optimization problem in mathematics which targets the NP-hard L0 problem. How to address the L0 sparsity problem is one of the major challenges in sparse unmixing. In literature [15], L1-norm based convex relaxation was used and the sparsity was represented by a regularizer. Following this work, improved approaches, such as L_p-norm ($0 < p \leq 2$) [16], [17], weighted $L_{d,1}$-norm ($d = 1, 2$) [18] and greedy algorithms [19], [20] were proposed. Spatial information is also widely used, which assumes that the types

Manuscript received June 7, 2020; revised July 20, 2020; accepted August 6, 2020. Date of publication August 26, 2020; date of current version March 25, 2021. This work was supported in part by the National Key Research and Development 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, in part by the China Post-Doctoral Science Foundation under Grant 2020M670631, and in part by the National Defense Science and Technology Innovation Special Zone Project. *(Corresponding author: Bin Pan.)*

Xia Xu and Tao Li are with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: xuxia@nankai.edu.cn; litao@nankai.edu.cn).

Bin Pan and Zongqing Chen are with the School of Statistics and Data Science, Nankai University, Tianjin 300071, China (e-mail: panbin@nankai.edu.cn; zqchern@nankai.edu.cn).

Zhenwei Shi is with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2020.3016941

of endmembers and the corresponding abundances in a single pixel are closely related to their neighboring ones. A variety of spatial information has been included for sparse unmixing, such as total variation [21], reweighted sparse [22], [23], region clustering [24], and multiscale decomposition [25]. Since sparse unmixing methods should rely on *a priori* spectral library, the spectral variations between the library and real hyperspectral data were considered [26]. Besides, the spectral variability within hyperspectral images were investigated [27], [28].

Recently, researchers have attempted to improve the performance of sparse unmixing by library pruning, which became the motivation of our work. Most sparse unmixing methods were designed under the basis of an overcomplete spectral library. Then there is a question: Whether the quality of the spectral library will affect the unmixing results. In recent studies, Iordache *et al.* [29]–[32] have pointed out that directly using the spectral library may make the sparse regression problem severely ill-conditioned. This is mainly because of the high mutual coherence of the signatures in the spectral library [29]. To address this challenge, Iordache *et al.* [31] proposed a multiple signal classification (MUSIC) and collaborative sparse regression (CSR)-based hyperspectral unmixing method called MUSIC-CSR. MUSIC-CSR is a two-stage approach, where it first uses MUSIC to prune the spectral library, and then conducts collaborative sparse unmixing. The authors further proved that the obtained spectral subset after MUSIC would include all the endmembers if there was no noise in the hyperspectral image. Based on this work, there are many improved works, such as [32]–[36]. MUSIC-CSR has indicated that library pruning is a promising manner to enhance the performance of sparse unmixing, which is the major motivation of our work.

However, existing library pruning algorithms may suffer endmember loss in high-noise conditions, since they adopted a two-stage manner. Although literature [31] has validated that the real endmembers will not be removed by MUSIC in noise-free hyperspectral imagery, it is nearly impossible to guarantee that no noise exists in real hyperspectral images. The situation will become even worse when the materials are highly mixed.

In this article, we propose a new multiobjective based one-stage algorithm which simultaneously conducts library pruning and sparse unmixing. The idea of MUSIC is reorganized as an optimization objective via an equivalent mathematical transformation, and this objective is further used to participate in a multiobjective optimization process. The proposed method is abbreviated as Pruning-based Multiobjective Sparse Unmixing (PMoSU). Multiobjective optimization refers to optimizing several objectives at the same time, rather than combining them into a single one. Recently, multiobjective methods have already been applied to hyperspectral unmixing [37]–[42]. In PMoSU, we design three objectives: reconstruction error, sparsity error, and signal projection error. PMoSU is not affected by the numerical values of objectives and thus there is no regularization coefficients required, and this is also a major reason why we select multiobjective optimization to construct our unmixing model. It seems that PMoSU has similar effects as adding a new regularizer. Unfortunately, adding a regularizer may bring in new problems: Many studies have demonstrated that the adjustment on regularization coefficients is quite an empirical work [18]–[21], [37]. Therefore, instead of adding a new regularizer, we transform the pruning and sparse representation to a multiobjective optimization problem and propose a simultaneous sparse unmixing and library pruning method to avoid the endmembers loss caused by pruning in high-noise conditions.

On the other hand, it is not feasible to directly adopt existing multiobjective framework to PMoSU. Multiobjective-based sparse unmixing methods may present a general problem: There are weakly Pareto optimums [43]. The weakly Pareto optimal problem may lead to nonuniqueness of the solutions, i.e., there are several solutions that are equally optimal. The reason is that one of the objectives in sparse unmixing, sparsity error, is discrete. As analyzed in literature [43], the discrete range is doomed to result in weakly Pareto optimal for the Tchebycheff-based decomposition approach. In this article, we consider the adaptive-penalty-based boundary intersection (APBI) approach which was provided in literature [43], and improve it to a three-direction extension. The improved APBI approach will guarantee that PMoSU can obtain the theoretically unique solution set. Overall, the major contributions of PMoSU can be summarized as follows.

1) We propose a pruning-based sparse unmixing method which can significantly reduce the mutual coherence of signatures in the spectral library. Compared with prepruning methods such as MUSIC-CSR, PMoSU can gradually compress the search space of sparse representation, while MUSIC-CSR directly compresses the search space, in which case, real endmembers may be missed when the noise is strong. Meanwhile, we have provided theoretical analysis and proof for the reasonability of the pruning strategy.
2) We adopt a multiobjective-based optimization approach so as to realize simultaneous pruning and unmixing, where the pruning projection is described as one of the objectives. Moreover, we further improve the APBI based decomposition during the optimization process to a three-direction extension, in which case the weakly Pareto optimal problem could be avoided and the theoretical unique solution set for sparse unmixing could be obtained.

This article is an extension for our conference version [44]. However, paper [44] only reports an idea, while detailed discussion and theoretical analysis are given in this article, especially the proofs for reasonability of the pruning strategy. Besides, more sufficient experiments are supplemented. Therefore, this article can be considered as a new and complete work.

II. Related Works

In this section, we will introduce the linear mixing model (LMM) and the definition of sparse unmixing, both of which are the basis of PMoSU.

A. LMM

Let $\mathbf{y} \in \mathbb{R}^{L\times 1}$ denote a pixel vector in a hyperspectral image with L dimensions. Then this pixel can be expressed by the following linear combination:

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

Equation (1) is the famous LMM, where $\mathbf{A} \in \mathbb{R}^{L\times m}$ is the endmember matrix with m materials, $\mathbf{x} \in \mathbb{R}^{m\times 1}$ is the abundance vector for all the m materials, and $\mathbf{n} \in \mathbb{R}^{L\times 1}$ is an additive perturbation including noise and modeling errors. Assume that there are n pixels in a hyperspectral image, then (1) can be organized as a matrix form

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_n]$ are the spectra, abundance, and noise matrices, respectively. In order not to violate the physical meaning of abundance, abundance nonnegativity constraint (ANC) and abundance sum-to-one constraint (ASC) are usually imposed into LMM [45]

$$\begin{aligned} &\text{ANC}: \ \mathbf{X} \geq 0 \\ &\text{ASC}: \ \forall \mathbf{x}_i \in \mathbf{X}, \quad \mathbf{1}^T\mathbf{x}_i = 1. \end{aligned} \tag{3}$$

Literature [15] has pointed out that ASC is a little too strong, meanwhile, ANC could give an automatic imposition to generalized ASC. So, in this article ASC is ignored.

B. Sparse Unmixing

Sparse unmixing is implemented under the basis of LMM. Sparse unmixing is a semisupervised process, where there is a prior spectral library including many real-world spectra available. To be specific, the endmember matrix $\mathbf{A}$ in LMM corresponds to the spectral library in sparse unmixing, and it is overcomplete. In this case, unmixing amounts to finding the optimal subset from the library that can best model the mixed pixels. Usually, the spectra number m in the library $\mathbf{A}$ is much larger than the real materials number. Therefore, the unmixing process can be written as a sparse representation problem

$$\begin{aligned} &\min_{\mathbf{X}\geq 0} \ \|\mathbf{X}\|_{\text{row}-0} \\ &\text{s.t.} \ \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\text{F}}^2 \leq \delta \end{aligned} \tag{4}$$

where $\|\mathbf{X}\|_{\text{row}-0}$ [46] is the number of nonzero rows in matrix $\mathbf{X}$, and $\delta \geq 0$ is the error tolerance due to noise and modeling errors.

Obviously, (4) is NP-hard. Usually, (4) is transformed to an L1-norm relaxation and a regularizer

$$\begin{aligned} &\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{\text{F}}^2 + \lambda \sum_{i=1}^{n} \|\mathbf{x}_i\|_1 \\ &\text{s.t.} \ \mathbf{X} \geq 0 \end{aligned} \tag{5}$$

where λ is a regularization coefficient that should be manually set. Some methods have also introduced new regularizers such as [21], [24], and [25], or used L_p-norm ($p < 1$) instead of L1 [16], [17]. Let $R(\mathbf{X})$ denote a certain regularizer, then the general sparse unmixing problem can be described by the following optimization form:

$$\begin{aligned} &\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda_1 \sum_{i=1}^{n} \|\mathbf{x}_i\|_p + \lambda_2 R(\mathbf{X}) \\ &\text{s.t.} \ \mathbf{X} \geq 0. \end{aligned} \tag{6}$$

However, the hyperparameters λ_1 and λ_2 have serious impact on the unmixing results. Users have to change λ_1 and λ_2 for different hyperspectral images. To some extent, the existence of λ_1 and λ_2 may make the sparse unmixing problem a little ill-posed. This situation becomes even worse when more regularizers are added. Furthermore, (6) is still a relaxation for the original L0-norm sparse problem, which may not obtain the real optimal solutions.

III. Methodology

In this section, we first introduce our multiobjective-based model. Then, the pruning operation is described by the signal projection error which acts as one of the objectives. Finally, we go into details about the multiobjective based optimization process, and our improvements are elaborated.

A. Multiobjective Based Model

To address the above problems, in this article we utilize a multiobjective based model for sparse unmixing. Instead of using (6), in this article we consider the reconstruction error, sparsity error, and the regularizer as three parallel objectives, and design the following multiobjective optimization problem:

$$\begin{aligned} \min_{\boldsymbol{s}\in\{0,1\}^m} F(\boldsymbol{s}) &= [f_1(\boldsymbol{s}), f_2(\boldsymbol{s}), f_3(\boldsymbol{s})]^{\text{T}} \\ f_1(\boldsymbol{s}) &= \|\mathbf{Y} - \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}}\|_{\text{F}}^2 \\ f_2(\boldsymbol{s}) &= \|\boldsymbol{s}\|_1 \\ f_3(\boldsymbol{s}) &= \|R(\mathbf{X}_{\boldsymbol{s}})\|_{\text{F}}^2. \end{aligned} \tag{7}$$

Equation (7) is not a simple transformation for (6), actually, it attempts to overcome the drawbacks of (6). The variable $\boldsymbol{s} = [s_1, \ldots, s_m]$ is a binary representation of the library $\mathbf{A}$, where $|\boldsymbol{s}| = \text{supp}(\mathbf{X})$. For hyperspectral unmixing problem, $s_i = 1$ denotes that the ith spectral signature is selected as an endmember and 0 otherwise. $\mathbf{A}_{\boldsymbol{s}}$ and $\mathbf{X}_{\boldsymbol{s}}$ are the subsets of $\mathbf{A}$ and $\mathbf{X}$, where only the positions with $s_i = 1$ remain. Different from (6), (7) is a subset selection problem in which case the solution contains all the selected endmembers. The corresponding abundances are calculated via nonnegative least squares. Note that $f_3(\cdot)$ can be any objective function, and in this article we will introduce pruning term. More details about $f_3(\cdot)$ will be shown in Section III-B.

According to literature [47], (7) is a multiobjective problem, where $f_1(\boldsymbol{s})$, $f_2(\boldsymbol{s})$ and $f_3(\boldsymbol{s})$ will be optimized simultaneously. Using the multiobjective framework will generate a significant advantage: There is no need to balance each objective and thus the setting of regularization coefficients λ_1 and λ_2 can be avoided. Furthermore, because we replace the decision variable from $\mathbf{X}$ to $\boldsymbol{s}$, it is possible to optimize L0 problem directly without any relaxation. Inspired by the above advantages, in this article we construct a multiobjective model which is denoted by (7).

B. Integrating Pruning and Unmixing

In the method MUSIC-CSR [31], MUSIC algorithm was used as preprocessing for sparse unmixing. MUSIC is a pruning strategy which can remove the redundancy spectra from the library. According to literature [31], MUSIC can weaken the high mutual coherence of the spectral library and thus improve the unmixing accuracy. However, MUSIC-CSR is a prepruning method. When the hyperspectral imagery includes noise, the real endmembers may miss after pruning [31].

In this article, we integrate the pruning operation into the process of optimization. Assume that the noise matrix $\mathbf{N} = 0$, then the objectives $f_1(\boldsymbol{s})$ and $f_2(\boldsymbol{s})$ can be reorganized as the following problem:

$$\begin{aligned} &\min \ \|\boldsymbol{s}\|_0 \\ &\text{s.t. } \mathbf{Y} = \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}}. \end{aligned} \tag{8}$$

In other words, we will target the problem described in (8) instead of the original sparse unmixing problem. Here, we give a theorem that (8) is equivalent to the original sparse unmixing problem, i.e., Theorem 1.

Theorem 1: The problem

$$\begin{aligned} &\min \ \|\mathbf{X}\|_0 \\ &\text{s.t. } \mathbf{Y} = \mathbf{A}\mathbf{X} \end{aligned} \tag{9}$$

is equivalent to the following problem:

$$\begin{aligned} &\min \|\boldsymbol{s}\|_0 \\ &\text{s.t. } \mathbf{Y} = \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}} \end{aligned} \tag{10}$$

where $\boldsymbol{s}$ is the subset of the column index set of $\mathbf{A}$, $\mathbf{A}_{\boldsymbol{s}}$ is a submatrix of $\mathbf{A}$ with columns indexed by $\boldsymbol{s}$, and $\mathbf{X}_{\boldsymbol{s}} \in \mathbb{R}^{|\boldsymbol{s}|\times n}$.

Proof: For the solution $\mathbf{X}$ of problem (9), let $\boldsymbol{s}$ be the set of indices of nonzero rows of $\mathbf{X}$. By interchanging rows, $\mathbf{X}$ can take the form

$$\begin{bmatrix} \mathbf{X}_{\boldsymbol{s}} \\ 0 \end{bmatrix} \tag{11}$$

where $\mathbf{X}_{\boldsymbol{s}}$ consists of all nonzero rows of $\mathbf{X}$. Also, $\mathbf{A}$ can take the form of $[\mathbf{A}_{\boldsymbol{s}}, \tilde{\mathbf{A}}]$ with the same column interchanging. Therefore

$$\begin{aligned} \mathbf{Y} &= \mathbf{A}\mathbf{X} \\ &= \left[\mathbf{A}_{\boldsymbol{s}}, \tilde{\mathbf{A}}\right]\begin{bmatrix} \mathbf{X}_{\boldsymbol{s}} \\ 0 \end{bmatrix} \\ &= \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}} + \tilde{\mathbf{A}} \cdot 0 \\ &= \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}}. \end{aligned} \tag{12}$$

Then, the corresponding $\boldsymbol{s}$ is the solution of (10), and vice versa. Since $\|\boldsymbol{s}\|_0 = \|\mathbf{X}\|_0$, according to the description in literature [48], (10) has a unique solution iff

$$\|\boldsymbol{s}\|_0 < \frac{\operatorname{spark}(\mathbf{A}) + \operatorname{rank}(\mathbf{Y}) - 1}{2} \tag{13}$$

where $\operatorname{spark}(\mathbf{A})$ is the smallest number of linearly dependent columns of $\mathbf{A}$. ∎

In hyperspectral unmixing problem, (13) can be satisfied by increasing $\operatorname{rank}(\mathbf{Y})$. According to (8) and (13), we can imply that $\operatorname{rank}(\mathbf{Y}) \le \|\boldsymbol{s}\|_0$ and $\|\boldsymbol{s}\|_0 \le \operatorname{spark}(\mathbf{A}) - 1$. This theorem also guarantees that our multiobjective strategy will not change the final target of sparse unmixing, and it only adjusts the optimization process.

Considering the physical meaning of $\mathbf{A}$ and $\|\boldsymbol{s}\|_0$, we can approximately simplify $\operatorname{rank}(\mathbf{Y}) \le \|\boldsymbol{s}\|_0$ to $\operatorname{rank}(\mathbf{Y}) = \|\boldsymbol{s}\|_0 = k$, where k is the number of linearly independent basis vectors of $\mathbf{Y}$, i.e., the number of endmembers. In this case, the support set $\operatorname{supp}(\mathbf{X})$ can be uniquely recovered as follows. If ignoring the noise $\mathbf{N}$, the hyperspectral data can be reconstructed by $\mathbf{Y} = \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}}$. Since $\mathbf{A}_{\boldsymbol{s}}$ and $\mathbf{X}_{\boldsymbol{s}}$ are both full rank matrices, we can get $\operatorname{range}(\mathbf{Y})$=$\operatorname{range}(\mathbf{A}_{\boldsymbol{s}})$. Therefore, the orthogonal basis of $\operatorname{range}(\mathbf{A}_{\boldsymbol{s}})$ can be replaced by the singular value decomposition for $\mathbf{Y}$.

Here, we use a projection to describe the pruning objective in PMoSU, which can be defined by the following theorem.

Theorem 2: Given the hyperspectral data $\mathbf{Y} = \mathbf{A}\mathbf{X}$, if it has the form $\mathbf{Y} = \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}}$ in the problem (10), where $\operatorname{rank}(\mathbf{Y}) = \operatorname{rank}(\mathbf{A}_{\boldsymbol{s}}) = \operatorname{rank}(\mathbf{X}_{\boldsymbol{s}}) = \|\boldsymbol{s}\|_0$. Then the column index $j \in \boldsymbol{s}$ iff $\mathbf{P}^{\perp}_{\mathbf{A}_{\boldsymbol{s}}}\mathbf{a}_j = 0$, where $\mathbf{P}^{\perp}_{\mathbf{A}_{\boldsymbol{s}}}$ is the projector on $\operatorname{Range}(\mathbf{A}_{\boldsymbol{s}})^{\perp}$.

Proof: By Autonee–Eckart–Young Theorem, $\mathbf{Y}$ has the following singular value decomposition:

$$\mathbf{Y} = U\Sigma V^T \tag{14}$$

where U and V are orthogonal matrices, and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$.

Write it as the following block matrix:

$$\begin{aligned} U &= \begin{bmatrix} U_k & \tilde{U} \end{bmatrix} \\ V &= \begin{bmatrix} V_k & \tilde{V} \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \Sigma_k & \\ & O_{n-k} \end{bmatrix} \\ \mathbf{Y} &= \begin{bmatrix} U_k & \tilde{U} \end{bmatrix}\begin{bmatrix} \Sigma_k & \\ & O_{n-k} \end{bmatrix}\begin{bmatrix} V_k^T \\ \tilde{V}^T \end{bmatrix} \\ &= U_k\Sigma_k V_k^T. \end{aligned} \tag{15}$$

Since $\Sigma_k\tilde{V}_k^T$ has full column rank, we have

$$\operatorname{Range}(\mathbf{Y}) = \operatorname{Range}(U_k). \tag{16}$$

By the orthogonality of U, we have $\operatorname{Range}(U_k) = \operatorname{Range}(\tilde{U})^{\perp}$, and

$$\begin{aligned} I &= UU^T \\ &= \begin{bmatrix} U_k & \tilde{U} \end{bmatrix}\begin{bmatrix} U_k^T \\ \tilde{U}^T \end{bmatrix} \\ &= U_kU_k^T + \tilde{U}\tilde{U}^T. \end{aligned} \tag{17}$$

Therefore, $\tilde{U}\tilde{U}^T = I - U_kU_k^T$.

By Theorem 1

$$\operatorname{Range}(\mathbf{Y}) = \operatorname{Range}(\mathbf{A}_{\boldsymbol{s}}). \tag{18}$$

It follows from (16) and (18) that $\operatorname{Range}(\mathbf{A}_{\boldsymbol{s}}) \perp \operatorname{Range}(\tilde{U})$.

And hence

$$\begin{aligned} j \in \boldsymbol{s} &\Leftrightarrow \mathbf{a}_j \perp \operatorname{Range}\left(\tilde{U}\right) \\ &\Leftrightarrow \tilde{U}^T\mathbf{a}_j = 0 \\ &\Leftrightarrow \tilde{U}\tilde{U}^T\mathbf{a}_j = 0. \end{aligned} \tag{19}$$

Let $P^{\perp}_{\mathbf{A}_{\boldsymbol{s}}} = \tilde{U}\tilde{U}^T$, then $j \in \boldsymbol{s} \Leftrightarrow P^{\perp}_{\mathbf{A}_{\boldsymbol{s}}}\mathbf{a}_j = 0$, equivalently, $P^{\perp}_{\mathbf{A}_{\boldsymbol{s}}}$ is the projector on $\operatorname{Range}(\mathbf{A}_{\boldsymbol{s}})^{\perp}$. ∎

Theorem 2 is the mathematical foundation of applying MUSIC to the hyperspectral unmixing problem, which indicates that the MUSIC-based pruning approach can be represented by a projection function in our multiobjective framework. Based on the above discussion, we can safely integrate Theorem 2 to the optimization process of PMoSU, and construct the following simultaneous optimization objectives:

$$\begin{aligned} \min_{\mathbf{s}\in\{0,1\}^m} F(\boldsymbol{s}) &= [f_1(\boldsymbol{s}), f_2(\boldsymbol{s}), f_3(\boldsymbol{s})]^{\mathrm{T}} \\ f_1(\boldsymbol{s}) &= \|\mathbf{Y} - \mathbf{A}_{\boldsymbol{s}}\mathbf{X}_{\boldsymbol{s}}\|_{\mathrm{F}}^2 \\ f_2(\boldsymbol{s}) &= \|\boldsymbol{s}\|_1 \\ f_3(\boldsymbol{s}) &= \|\mathbf{P}_{\mathbf{A}_{\boldsymbol{s}}}^{\perp}\mathbf{A}_{\boldsymbol{s}}\|_{\mathrm{F}}^2 \end{aligned} \tag{20}$$

where $\mathbf{A}_{\boldsymbol{s}} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$, k is the number of selected spectral signatures. $\|\mathbf{P}_{\mathbf{A}_{\boldsymbol{s}}}^{\perp}\mathbf{A}_{\boldsymbol{s}}\|_{\mathrm{F}}^2 = \sum_{i=1}^{\|\boldsymbol{s}\|_1}\|\mathbf{P}_{\mathbf{A}_{\boldsymbol{s}}}^{\perp}\mathbf{a}_j\|^2$ is the sum of projections.

There are two major differences between the proposed method and MUSIC-CSR [31]. One difference is that $f_3(\boldsymbol{s})$ takes all selected spectral signatures into consideration, so that we do not need to calculate the projection one by one. Better spectral signature corresponds to a smaller projection $\|\mathbf{P}_{\mathbf{A}_{\boldsymbol{s}}}^{\perp}\mathbf{a}_j\|^2$, thus better spectral signatures combination may have a smaller $\|\mathbf{P}_{\mathbf{A}_{\boldsymbol{s}}}^{\perp}\mathbf{A}_{\boldsymbol{s}}\|_{\mathrm{F}}^2$. Another difference is that our method keeps updating the selected $\mathbf{A}_{\boldsymbol{s}}$ during the optimization process, and the final determined $\mathbf{A}_{\boldsymbol{s}}$ is the very solution. By comparison, MUSIC-CSR is a prepruning operation. In fact, the obtained $\mathbf{A}_{\boldsymbol{s}}$ in MUSIC-CSR is still overcompleted, and users have to further conduct sparse unmixing by considering $\mathbf{A}_{\boldsymbol{s}}$ as a new library. To some extent, our method is a relaxation for MUSIC-CSR. According to Theorem 1, our transferred subset selection problem is equivalent to the original sparse unmixing problem. By comparison, MUSIC-CSR actually has changed the original form of sparse unmixing, since the dictionary is only a subset of the original one. Although we have ignored the noise in the derivation of Theorem 1, the endmembers loss in PMoSU can be considered as random error, while the endmembers loss in prepruning methods is systematic error. When the hyperspectral data are noisy, $f_1(\boldsymbol{s})$, $f_2(\boldsymbol{s})$ and $f_3(\boldsymbol{s})$ work together to generate better reconstruction results as well as avoid pruning the real endmembers by mistake.

C. APBI-Based Optimization

We solve (20) using the framework of multiobjective evolutionary algorithm based on decomposition (MOEA/D) [49]. Because the range space of (20) is discrete which may result in weakly Pareto optimal, in PMoSU we use APBI to improve MOEA/D. Fig. 1 is a simplified illustration for the optimization process of PMoSU. We will explain the terms in Fig. 1 during the introduction to the algorithm steps.

The optimization process is iteratively conducted, which mainly includes five steps: initialization; new individuals generation; new population determination; Pareto front (PF) update; stopping criteria. Among them how to determine new population is the key point, and our improvements are also conducted on this step.

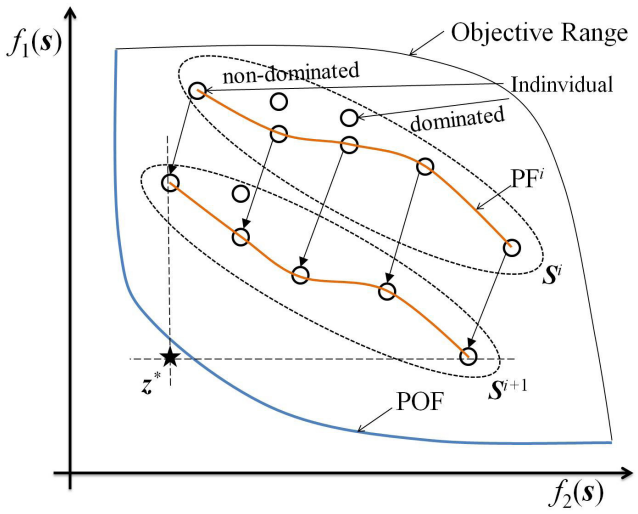


Fig. 1. Simplified illustration for the optimization process of PMoSU. To make the figure clearer, here we only use two objectives for example.

1) Initialization: PMoSU begins with an initial solution set $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p\}$. Please note that multiobjective methods usually generate a solution set rather than a single solution. The set $\boldsymbol{S}$ is also called *population* during the optimization process, and its size is p. Each element in $\boldsymbol{S}$ corresponds to a solution $\boldsymbol{s}$ for (20). During the optimization process, the element $\boldsymbol{s}$ is called *individual*. In PMoSU, we set $\boldsymbol{S}$ by zero vectors for initialization. Individuals in $\boldsymbol{S}$ will be diverse after updating.

2) New Individuals Generation: Bitwise flipping strategy [38] is conducted for all the individuals among current population to generate new individuals. For one individual $\boldsymbol{s}_i$ which is a binary vector, bitwise flipping refers to flipping each element from 0 to 1 (or 1 to 0) with a probability $1/m$. After flipping, the individual $\boldsymbol{s}_i$ is changed to $\boldsymbol{s}_i'$. Conducting this strategy on all in individuals in $\boldsymbol{S}$, we can get a new set $\boldsymbol{S}'$.

3) Principle for Population Update: Here, we firstly introduce two terms, *dominate* and *ideal point*. Assume $\boldsymbol{s}_u$ and $\boldsymbol{s}_v$ are two solutions for (20), we say $\boldsymbol{s}_u$ dominates $\boldsymbol{s}_v$ iff

$$\forall i \in \{1, 2, 3\}, \quad f_i(\boldsymbol{s}_u) < f_i(\boldsymbol{s}_v). \tag{21}$$

A natural idea to determine the new population is finding all the optimal nondominated solutions among the union set $\{\boldsymbol{S}, \boldsymbol{S}'\}$. Actually this is also the idea used by early multi-objective methods such as NSGA-II [50]. However, these methods were usually time consuming, because users have to conduct pairwise comparison on $\{\boldsymbol{S}, \boldsymbol{S}'\}$. For example, assume $\boldsymbol{S}$ and $\boldsymbol{S}'$ each have 100 individuals, then there should be C_{100}^2 comparisons.

To overcome the efficiency problem, decomposition based multiobjective methods were proposed, and MOEA/D is one of the representatives. MOEA/D tries to update the population by comparing the distance between an *ideal point* and the individuals before and after flipping. In PMoSU, the ideal point (noted by $\boldsymbol{z}^*$) is a virtual point which is calculated by

the following equation:

$$z^* = \left(z_1^*, z_2^*, z_3^*\right)^{\mathrm{T}}$$
$$\text{where } z_i^* = \min\{f_i(\boldsymbol{s}) | \boldsymbol{s} \in \boldsymbol{S}\}. \tag{22}$$

z^* is a virtual point which is updated along with the population during each iteration. PMoSU tries to force the solution set to approach the ideal point using an iteration manner.

4) Population Update: In MOEA/D, the distance between an individual $\boldsymbol{s}_i$ and the ideal point z^* is calculated by weighted Tchebycheff decomposition

$$g_i^{\mathrm{wt}}\left(\boldsymbol{s}_i | \boldsymbol{\lambda}_i, z^*\right) = \max_{1 \le j \le 3} \left\{\lambda_i^j | f_j(\boldsymbol{s}_i) - z^*|\right\} \tag{23}$$

where $\boldsymbol{\lambda}_i = [\lambda_1, \lambda_2, \lambda_3]^{\mathrm{T}}$ is a direction vector for the ith individual. $\boldsymbol{\lambda}_i$ is a uniformly distributed direction vector with $(\lambda_1 + \lambda_2 + \lambda_3 = 1)$. The usage of $\boldsymbol{\lambda}_i$ can partly avoid the influence of data scalar. For more detailed description about weighted Tchebycheff decomposition, refer to literature [49].

For all the individuals in $\boldsymbol{S}$ and $\boldsymbol{S}'$, (23) can provide their distances to z^*. Since weighted Tchebycheff decomposition is a direction-based distance measure, users only require to compare the distances between $\boldsymbol{s}_i$ and $\boldsymbol{s}_i'$ for the ith individual. Therefore, if we still assume there are 100 individuals, MOEA/D only needs 100 comparisons, which is much less than NSGA-II. This is the major reason why MOEA/D is much more efficient than NSGA-II, which makes MOEA/D become more and more popular in recent years. In practice, users usually not only compare the optimality between $\boldsymbol{s}_i$ and $\boldsymbol{s}_i'$, but also slightly expand the comparison scope to the neighbors around $\boldsymbol{s}_i$. Let $\boldsymbol{B}_i$ denote a subset of $\boldsymbol{S}$, which contains several closest individuals around $\boldsymbol{s}_i$. During each iteration, MOEA/D updates the ith individual by minimizing the Tchebycheff distances between z^* and $\{\boldsymbol{s}_i, \boldsymbol{s}_i', \boldsymbol{B}_i\}$. The number of individuals in $\boldsymbol{B}_i$ is a hyperparameter, which is set as 20 in PMoSU.

However, the solution range of PMoSU is discrete, as shown in (20), $f_2(\boldsymbol{s})$. For weighted Tchebycheff decomposition-based multiobjective methods, the discrete range may result in weakly Pareto optimal. In this article, we have improved the weighted Tchebycheff decomposition approach by APBI [43].

APBI is a new distance measure aiming at weakly Pareto optimal problem. According to the conclusion of [43], the distance between an individual $\boldsymbol{s}_i$ and the ideal point can be obtained by

$$g_i^{\mathrm{a}}\left(\boldsymbol{s}_i | \boldsymbol{\lambda}_i, z^*\right) = d_1 + \theta\left(\arctan\left(\frac{\lambda_i^1}{\lambda_i^2}\right)\right) d_2 \tag{24}$$

where

$$d_1 = \frac{\|(\mathcal{F}(\boldsymbol{s}_i) - z^*)\boldsymbol{\lambda}_i\|_{\mathrm{F}}}{\|\boldsymbol{\lambda}_i\|_{\mathrm{F}}}$$
$$d_2 = \|\mathcal{F}(\boldsymbol{s}_i) - \left(z^* + d_1 \boldsymbol{\lambda}_i^{\mathrm{T}}\right)\|_{\mathrm{F}}$$
$$\theta\left(\arctan\left(\frac{\lambda_i^1}{\lambda_i^2}\right)\right) = 0.47\left(\arctan\left(\frac{\lambda_i^1}{\lambda_i^2}\right) - 0.78\right)^2 + 0.7 \tag{25}$$

$\theta(\cdot)$ is an adaptive penalty function. Different from literature [43], in PMoSU there are three objectives, in which case $\boldsymbol{\lambda}_i = [\lambda_i^1, \lambda_i^2, \lambda_i^3]^{\mathrm{T}}$ includes three elements. Therefore, we improve (26) to a three-direction form

$$g_i^{\mathrm{p}}\left(\boldsymbol{s}_i | \boldsymbol{\lambda}_i, z^*\right) = d_1 + \theta\left(\arctan\left(\frac{\lambda_i^1}{\|\left[\lambda_i^1, \lambda_i^2\right]^{\mathrm{T}}\|_{\mathrm{F}}}\right)\right) d_2. \tag{26}$$

The distance measure in PMoSU is finally calculated by $g_i^{\mathrm{p}}(\boldsymbol{b}_i | \boldsymbol{\lambda}_i, z^*)$.

After comparing $g_i^{\mathrm{p}}(\boldsymbol{s}_i | \boldsymbol{\lambda}_i, z^*)$ and $g_i^{\mathrm{p}}(\boldsymbol{s}_i' | \boldsymbol{\lambda}_i, z^*)$, we can determine which one is closer to z^*. The closer individual is collected to the new population, and the farther one is abandoned. Conducting this process for all the pairs in $\mathbf{S}/\mathbf{S}'$, a new population will appear.

5) PF Update and Stop Criteria: We first introduce the definition of PF. PF refers to function values of all the optimal *nondominated* solutions in current population. In each iteration, the PF is updated along with the population. With the increase of iterations, the PF will tend to the Pareto optimal front (POF). POF is the global optimal objective values, and in PMoSU we try to force PF as close to POF as possible.

The optimization process will stop as long as the solution set keeps stable among several iterations. To be specific, the optimization process will stop if the solution set is not changed within ten iterations. We consider that current PF is approximately equal to POF. In this case, the final solution set is obtained. Please note that after sufficient iterations we will get a solution set $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p\}$ rather than a single solution. Although $\boldsymbol{S}$ contains p solutions, some of them are

Algorithm 1: Pseudocode for PMoSU

Input: hyperspectral image data $\mathbf{Y}$, spectral library $\mathbf{A}$.
Output: abundance fractions $\mathbf{X}$.
Initialization:
population size p, maximum iteration number T, a population $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p\}$, a set of weight vector $\Lambda = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p\}$, the ideal point $\mathbf{z}^*$.
Endmember Selection:
while $t < T$ **do**
  $t = t + 1$;
  for $i = 1, \ldots, p$ **do**
    Generate a new individual $\boldsymbol{s}_i'$ from $\boldsymbol{s}_i$ based on bit-wise flipping strategy, where each location is flipped with a probability $1/m$;
    if $\|F(\boldsymbol{s}^*)\|_2 > \|F(\boldsymbol{s}_i')\|_2$ **then**
      Set $z^* = F(\boldsymbol{s}_i')$
    for $j \in \boldsymbol{B}_i$ **do**
      if $g_i^p(\boldsymbol{s}_i' | \boldsymbol{\lambda}_i, \boldsymbol{z}^*, \boldsymbol{s}^*) \le g_j^p(\boldsymbol{s}_j | \boldsymbol{\lambda}_j, \boldsymbol{z}^*, \boldsymbol{s}^*)$ **then**
        Set $\boldsymbol{s}_j = \boldsymbol{s}_i'$ and $F(\boldsymbol{s}_j) = F(\boldsymbol{s}_i')$
Return the final solution as $\boldsymbol{s}^\star$ and record the corresponding spectral signatures.
Abundance Estimation:
Compute the abundances for the whole hyperspectral image based on nonnegative least squares algorithm: $\mathbf{X} = \arg\min_{\mathbf{X} \ge 0} \|\mathbf{Y} - \mathbf{A}_{\boldsymbol{s}^\star}\mathbf{X}\|_{\mathrm{F}}$

overlapped. Therefore, the final number of solutions is less than p.

So far, all the terms in Fig. 1 have been defined. Then the overall flow of PMoSU can be described as follows. First, initialize a population $\boldsymbol{S}_0$ and calculate an ideal point. Second, conduct bitwise flipping for all the individuals in $\boldsymbol{S}_0$, and thus generate new population $\boldsymbol{S}_0'$. Then, compare the APBI distance among individuals in $\{\boldsymbol{S}_0, \boldsymbol{S}_0'\}$, and collect all the nearest individuals to construct $\boldsymbol{S}_1$ as well as new ideal point. Run the above process iteratively until the population keep stable. Current population are considered as the final solution set. The pseudocode of PMoSU is shown in Algorithm 1.

Similar to general multiobjective methods, the final solution is selected from the solution set by the knee point [37], [38], [40]. The knee point is one of the solutions on PF for which any improvement in one objective will result in a severe degradation in at least another one [51], [52]. Based on the obtained solution $\boldsymbol{s}$, the final endmember matrix $\mathbf{A_s}$ can be determined. The final abundance maps $\mathbf{X_s}$ can be solved by nonnegative constrained least squares (NCLS).

IV. Experiments and Discussion

Two synthetic data and one real hyperspectral image are used to validate the effectiveness of PMoSU. Our experiments mainly include three aspects.

1) Unmixing accuracies comparison among PMoSU and other state-of-the-art sparse unmixing methods.
2) The discussion on the pruning process. Since PMoSU is a one-stage simultaneous pruning and classification algorithm which is improved from the two-stage method MUSIC-CSR, we should verify that PMoSU can avoid endmembers missing.
3) We will validate that the improvement by APBI can eliminate the influence of weakly Pareto optimal solutions.

We compare PMoSU with seven recently proposed sparse unmixing methods, namely SUnSAL [15], SUnSAL-TV [21], MUSIC-CSR [31], RSFoBa [19], SMP [20], SMoSU [41], and CM-MoSU [42]. SUnSAL was a classical sparse unmixing method which adopted convex relaxation to address the L0 problem. SUnSAL-TV, RSFoBa, and SMP have improved SUnSAL by adding new regularizers. RSFoBa and SMP adopted greedy algorithms to solve the sparse unmixing problem. SMoSU and CM-MoSU are recently proposed multiobjective based sparse unmixing methods which mainly focused on the optimization process. MUSIC-CSR is our baseline where pruning was used as a preprocessing approach.

The unmixing accuracies are evaluated by signal-to-reconstruction error (SRE) for the estimated abundance maps, which is defined by

$$\mathrm{SRE} = 10\lg\left(\mathrm{E}\left[\|\mathbf{X}\|_{\mathrm{F}}^2\right]/\mathrm{E}\left[\|\mathbf{X}-\hat{\mathbf{X}}\|_{\mathrm{F}}^2\right]\right) \tag{27}$$

where $\hat{\mathbf{X}}$ is the estimated abundance by algorithms. SRE is a popular criterion for unmixing problem, and it is also used in quite a few literatures [15], [16], [25].

Fig. 2. Subscene of the AVIRIS Cuprite data set.

A. Synthetic and Real Data

Because it is almost impossible to obtain the ground truth of real hyperspectral data, researchers usually use synthetic data for the quantitative analysis in hyperspectral unmixing task. In this article, we construct two synthetic data by different generation manners. Since the target of PMoSU is to avoid the endmember loss by MUSIC in high-noise conditions, we mainly discuss the comparison results with 20-/25-/30-dB noise. To be specific, we manually add 20-/25-/30-dB correlated noise to all the synthetic data, and the noise intensity is defined by signal-to-noise ratios (SNRs) (SNR $= 10\lg(\|\mathbf{Y}\|_{\mathrm{F}}^2/\|\mathbf{N}\|_{\mathrm{F}}^2)$). Besides, we use a real hyperspectral image for qualitative illustration. The experiments on real data aim at validating the effectiveness of PMoSU, rather than its superiority. Please note that all the synthetic data generation manners are widely used by many sparse unmixing works [36]–[38], [40]–[42].

1) Generation of Synthetic Data: The synthetic data are randomly constructed using the spectra of the United States Geological Survey (USGS) digital spectral library (splib06a).[1] Splib06a is composed of 498 spectral signatures with 224 spectral bands. Specially, splib06a is also used as the library of all the compared methods, i.e., the matrix $\mathbf{A}$ in (6). We select ten spectra from splib06a to construct the synthetic data. To increase the unmixing difficulty, five similar spectra are selected, namely Actinolite HS116.3B, Actinolite HS22.3B, Actinolite HS315.4B, Actinolite NMNH80714 and Actinolite NMNHR16485, and the remaining five are randomly selected. We validate the unmixing methods with endmember numbers varying from 4 to 10.

Synthetic data 1 contain 64×64 pixels whose abundance maps are Dirichlet distribution. Synthetic data 2 are generated using the same approach as literature SUnSAL-TV [21], where pixels among them are spatially dependent. Literature [21] provided only the five-endmember synthetic data generation approach. To better validate the influence of endmember numbers, in PMoSU we extend the original data generation approach to any number of endmembers. The major difference between synthetic data 1 and 2 is that the pixels in the latter

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

TABLE I

SRE VALUES (dB) BY DIFFERENT METHODS ON SYNTHETIC DATA 1 AND 2 WITH 20-dB CORRELATED NOISE. THE SECOND COLUMN CORRESPONDS TO THE NUMBER OF ENDMEMBERS THAT ARE USED TO GENERATE THE SYNTHETIC DATA

Data	Num.	SUnSAL	SUnSAL-TV	MUSIC-CSR	SMP	RSFoBa	SMoSU	CM-MoSU	PMoSU
Synthetic 1	4	3.286	2.713	6.123	2.524	7.922	9.376	10.02	**13.05**
	5	3.018	2.333	4.627	3.284	8.561	**8.859**	**8.859**	8.817
	6	3.363	2.878	6.712	5.701	5.144	5.096	9.122	**10.22**
	7	3.252	2.936	4.103	4.000	3.805	6.145	8.004	**8.111**
	8	2.382	2.227	0.488	3.973	4.627	5.485	**7.705**	6.098
	9	2.760	2.667	2.420	3.893	3.825	6.604	8.493	**8.577**
	10	3.104	2.888	1.335	4.712	3.256	5.562	9.412	**9.650**
Synthetic 2	4	5.927	9.530	6.058	7.469	7.915	10.08	**11.21**	10.42
	5	5.174	9.362	5.183	7.029	4.892	9.098	**10.76**	**10.76**
	6	5.224	**8.787**	4.995	6.774	3.445	7.735	7.597	7.735
	7	4.638	7.878	5.063	5.417	1.137	5.856	6.788	**7.880**
	8	3.895	6.992	4.704	5.207	3.105	6.663	6.663	**7.069**
	9	3.972	**7.092**	5.973	2.644	1.288	3.523	5.268	6.364
	10	2.824	**6.624**	0.063	3.642	1.546	4.399	4.631	5.411

TABLE II

SRE VALUES (dB) BY DIFFERENT METHODS ON SYNTHETIC DATA 1 AND 2 WITH 25-dB CORRELATED NOISE. THE SECOND COLUMN CORRESPONDS TO THE NUMBER OF ENDMEMBERS THAT ARE USED TO GENERATE THE SYNTHETIC DATA

Data	Num.	SUnSAL	SUnSAL-TV	MUSIC-CSR	SMP	RSFoBa	SMoSU	CM-MoSU	PMoSU
Synthetic 1	4	4.007	5.124	11.33	8.292	9.334	12.15	12.35	**13.24**
	5	2.774	4.354	3.382	3.710	3.738	11.71	11.71	**14.60**
	6	2.259	4.965	7.265	1.961	2.999	10.83	11.04	**13.11**
	7	2.875	5.235	4.525	3.643	7.367	9.485	10.91	**12.39**
	8	3.123	3.276	5.237	4.534	4.080	9.652	10.22	**11.58**
	9	2.367	2.617	9.051	7.108	2.906	6.756	11.25	**12.58**
	10	2.057	2.459	3.115	1.627	2.437	6.387	**11.88**	10.44
Synthetic 2	4	8.806	11.84	9.428	14.99	**15.76**	14.26	13.29	13.63
	5	7.473	11.40	8.240	11.95	13.19	12.17	12.17	**13.57**
	6	7.530	11.60	8.244	11.92	11.47	12.24	12.43	**13.17**
	7	6.940	10.33	7.849	10.62	8.677	11.19	**11.39**	11.20
	8	6.261	9.415	7.527	8.886	7.645	9.211	9.312	**9.703**
	9	6.423	9.193	8.172	8.090	9.906	7.178	**10.64**	9.906
	10	5.656	9.174	7.545	7.330	8.773	8.439	9.035	**9.370**

have spatial correlations, while the former do not. We force all the abundances smaller than 0.7 so as to make the unmixing process more challenging.

2) Real Data: A subscene of the AVIRIS Cuprite data set[2] is used for comparison. As a practice, several water absorption and noisy bands (bands 1 and 2, 105–115, 150–170, and 223 and 224) are removed, and 188 bands remain. A false-color composite image of this data is shown in Fig. 2.

B. Comparison of Unmixing Accuracies

Tables I–III display the quantitative evaluation results on the two synthetic data which are generated by different endmember numbers. We consider that Table II is the most meaningful among the three tables, which reports the results on 25-dB noise. Therefore, we first focus on Table II. It can be observed that PMoSU outperforms other methods in most cases. Generally, with the increasing of real endmembers numbers, the accuracies of all the methods tend to decline. This is because synthetic data with more endmembers can better simulate the highly mixed situation. It is observed that most of the compared methods have presented significant declines when spectra are highly mixed, and this situation is especially serious on synthetic data 1. PMoSU has presented 3–4-dB SREs gaps between endmember numbers 4 and 10. CM-MoSU and SUnSAL-TV decline less, which indicates that they may not be sensitive to the change of endmembers. However, when the endmember number is fixed, PMoSU achieves higher SREs. Specially, the robustness of PMoSU can be inferred from the comparison with MUSIC-CSR. In synthetic data 1, MUSIC-CSR achieves 11.33-dB SRE value when endmember number is 4, but it decreases to 3.115 dB when endmember

[2]Available online: http://lesun.weebly.com/hyperspectral-data-set.html

TABLE III

SRE VALUES (dB) BY DIFFERENT METHODS ON SYNTHETIC DATA 1 AND 2 WITH 30-dB CORRELATED NOISE. THE SECOND COLUMN CORRESPONDS TO THE NUMBER OF ENDMEMBERS THAT ARE USED TO GENERATE THE SYNTHETIC DATA

Data	Num.	SUnSAL	SUnSAL-TV	MUSIC-CSR	SMP	RSFoBa	SMoSU	CM-MoSU	PMoSU
Synthetic 1	4	4.139	8.963	12.87	15.26	10.26	**25.10**	**25.10**	**25.10**
	5	3.610	7.667	9.360	8.549	5.899	**31.92**	**31.92**	**31.92**
	6	2.474	7.919	7.409	4.197	4.139	21.89	**22.36**	**22.36**
	7	4.116	7.850	6.469	5.484	7.428	**16.36**	**16.36**	**16.36**
	8	2.462	7.190	4.321	4.239	6.076	17.34	**18.00**	17.89
	9	3.344	7.522	6.177	8.819	4.308	**20.40**	**20.40**	**20.40**
	10	4.573	8.095	8.480	6.827	4.919	**15.09**	**15.09**	**15.09**
Synthetic 2	4	10.73	16.39	14.08	17.81	16.85	18.77	**20.72**	**20.72**
	5	9.326	15.27	12.73	16.72	14.25	16.81	**18.05**	15.70
	6	9.359	15.25	11.89	16.25	13.81	16.58	16.52	**17.43**
	7	8.599	13.56	10.76	10.41	12.77	14.40	**15.36**	14.92
	8	7.859	11.90	10.06	11.26	11.51	**13.88**	**13.88**	**13.88**
	9	8.057	11.82	10.37	10.69	12.29	**13.61**	**13.61**	**13.61**
	10	7.662	11.42	9.989	5.826	11.40	10.62	11.35	**12.26**

number increases to 10. The reason for this phenomenon is that MUSIC algorithm has missed some endmembers during the pruning process, which is the very issue that PMoSU hopes to solve.

Table I shows the different unmixing results on 20-dB noise. Unfortunately, 20 dB is too strong for most algorithms. Both MUSIC-CSR and PMoSU have observed missing endmembers. As is discussed above, endmember missing will result in larger SRE decline than endmember redundancy. However, in addition to these endmembers missing situations, PMoSU performs well in most cases. Specially, PMoSU outperforms MUSIC-CSR by a large margin, which may indicate that the endmembers missing in PMoSU is less serious than that in MUSIC-CSR. Compared with SMoSU and CM-MoSU, the proposed simultaneous pruning strategy has presented significant advantages, similar to the results of 25-dB experiments. On the other hand, although SUnSAL-TV performs well on synthetic data 2, the selection of hyperparameters is difficult. Actually, different hyperparameters will significantly affect the accuracies of SUnSAL-TV, which may harm the robustness of SUnSAL-TV. By comparison, hyperparameters have little influence on PMoSU, because the most important ones, regularization coefficients, do not exist in PMoSU.

Table III reports the unmixing results by different algorithms on 30dB noise. We note that in Table III CM-MoSU and PMoSU have presented comparative accuracies which are slightly better than SMoSU, and all of them performs better than others. Actually, in most cases, CM-MoSU and PMoSU can find the exact endmembers neither less or more. Since they are all constructed under the basis of MOEA/D framework, they may make similar mistakes on the same condition. Generally, with the increase of endmember numbers, the performance of most unmixing methods decline. However, the significance of Table III seems not apparent, because previous methods such as CM-MoSU have presented nearly exact results. Therefore, we can safely conclude that 30 dB is not a strong noise for multiobjective based unmixing methods.

Fig. 3 is an illustration for the abundance maps obtained by different methods. Here, we take synthetic data 2 with five endmembers and 25-dB noise for example. We can find the noise in this data is strong, which may harm the unmixing results. The maps by SMoSU and CM-MoSU are the same, because they have extracted the same endmembers. Because SunSAL-TV has considered the spatial information, its abundance maps look clearer, but the SRE can be further improved. By comparison, PMoSU slightly outperforms others.

The experimental results on real data are shown in Fig. 4. Please note that in hyperspectral unmixing problem, it is almost impossible to obtain the abundances ground truth of different materials in real data. Besides, some popular criterions such as reconstruction error and the average spectral angle distance are affected by the selection of objective functions, which may not accurately describe the superiority of different algorithms. Therefore, real data are usually used for qualitative analysis, and they cannot be used to evaluate which unmixing method is quantitatively better. According to Fig. 4, most methods perform well. Note that in Fig. 4(c) MUSIC-CSR appears an endmember missing, which may be caused by the complex mixing condition in real hyperspectral data. It is difficult to judge which method is superior, but we can safely conclude that all of them are effective.

C. Analysis for Pruning and Weakly Pareto Optimal

This section includes two aspects: 1) the discussion about pruning process and 2) the discussion about weakly Pareto optimal.

Firstly, we compare PMoSU with two simplified versions, MoSU and MMoSU. MoSU only uses two objectives, $f_1(\cdot)$ and $f_2(\cdot)$, and MMoSU conducts MUSIC algorithm before MoSU. All the settings in MoSU, MMoSU, and PMoSU keep the same. Note that MMoSU is different from MUSIC-CSR. Both MoSU and MMoSU are modified from PMoSU. We design this experiment to validate the effectiveness of the simultaneous pruning strategy.

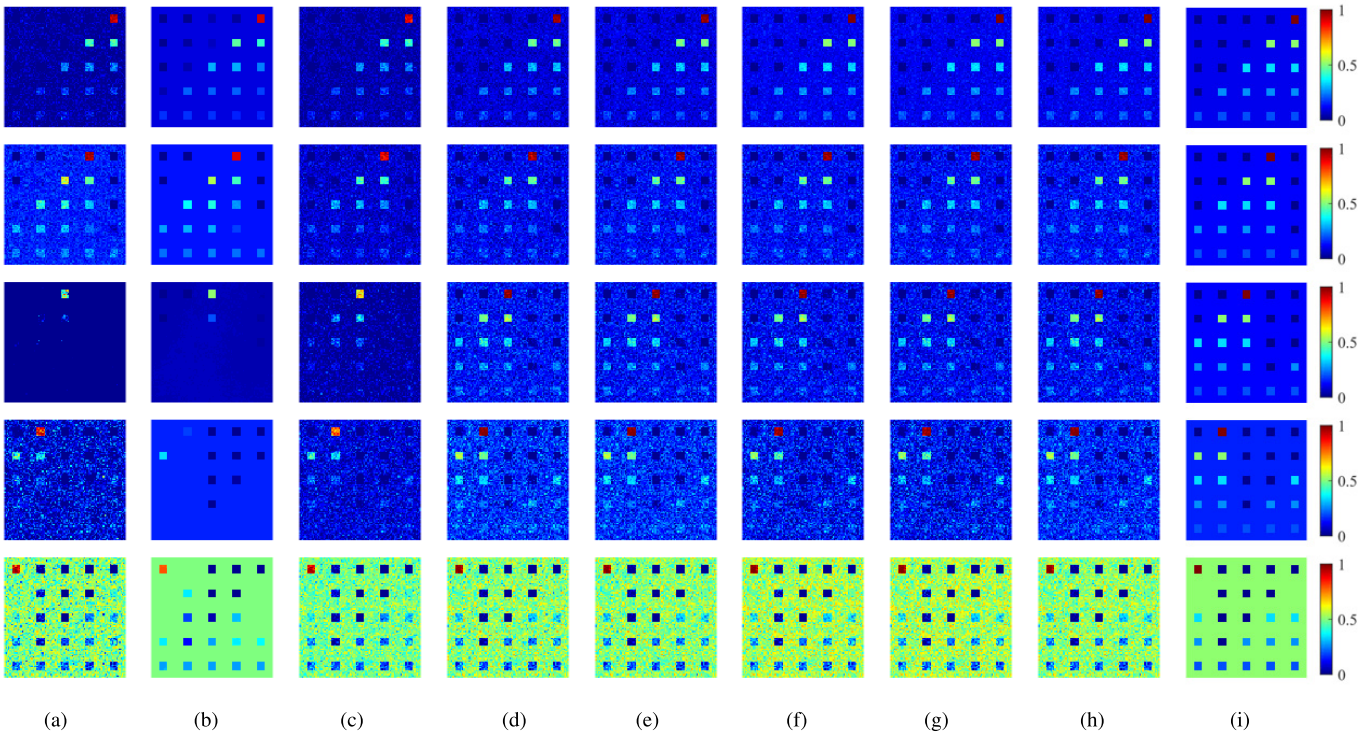


Fig. 3. Abundance maps on synthetic data 2 with five endmembers and 25-dB noise. Abundance maps obtained by (a) SUnSAL, (b) SUnSAL-TV, (c) MUSIC-CSR, (d) SMP, (e) RSFoBa, (f) SMoSU, (g) CM-MoSU, (h) PMoSU and (i) ground truth. (From top to bottom) Maps corresponding to endmembers #1–#5.

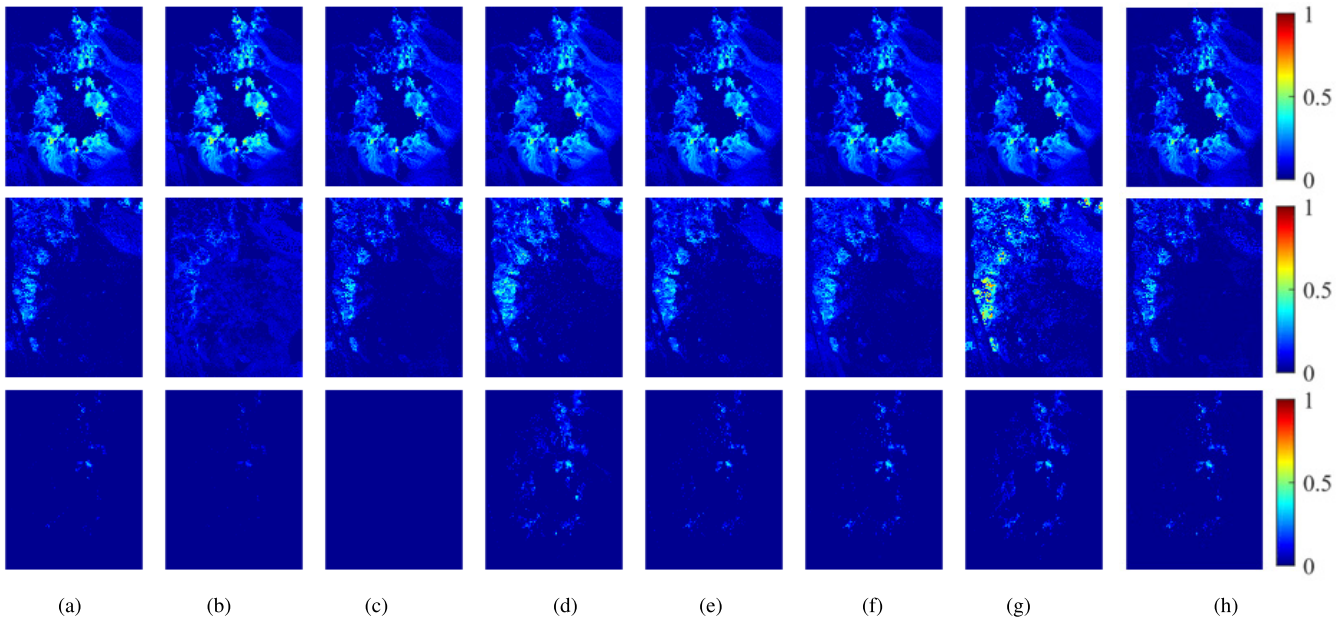


Fig. 4. Comparison of abundance maps on Cuprite data. Abundance maps obtained by (a) SUnSAL, (b) SUnSAL-TV, (c) MUSIC-CSR, (d) SMP, (e) RSFoBa, (f) SMoSU, (g) CM-MoSU, and (h) PMoSU. (From top to bottom) Maps corresponding to Alunite+Muscovite/Kaolinite, Hematite, and Alunite.

We take synthetic data 2 as an example, and the results are shown in Table IV. This table displays the selected endmembers' serial numbers by different methods. We number the 1st to the 498th spectra of the library in order. For simplicity, we set spectra at the front as the real endmembers that are used to generate synthetic data. For example, "Num.=4" means the first to fourth spectra are used in synthetic data, and the [1st, 2nd, 3rd, 4th, 118th and 170th] spectra are selected as the endmembers by PMoSU. The optimization process will stop if the solutions are not changed within ten iterations.

It is observed that both MMoSU and PMoSU perform better than MoSU in most cases, which demonstrates that the pruning strategy works well. MMoSU seems to achieve more precise results than PMoSU, especially when the endmembers

TABLE IV

SELECTED ENDMEMBERS' SERIAL NUMBERS BY DIFFERENT METHODS ON SYNTHETIC DATA 2, 25-dB NOISE. THE COLUMN "TIME" REFERS TO THE REQUIRED ITERATION TIME UNTIL THE SOLUTIONS KEEP STABLE, AND IT IS DENOTED BY MoSU/MMoSU/PMoSU

Num.	MoSU	MMoSU	PMoSU	Time(s)
4	1,2,3,4, 87,235	1,2,3,4, 18,44	1,2,3,4, 118,170	100/5.3/23.9
5	2,3,5,29, 54,195,218	1,2,3,4, 5,13,31	1,2,3,4,5,30, 110,170,186	60.7/22.4/138
6	1,2,4,6, 197,217,230	1,2,3,4, 5,6,10	1,2,3,4,5,6, 19,53,115,170	98.4/60.3/129
7	1,2,5,36, 161,192,234	1,2,3,4, 5,6,7	1,2,3,4,5, 6,7,19,68	155/14.5/180
8	1,4,5,49, 186,200,240	1,2,3,4, 5,6,7	1,2,3,4,5,6, 7,8,103,125	115/28.9/117
9	1,4,5,6, 9,42,95	1,2,3,4, 5,7,9	1,2,3,4,5,6, 7,8,36,170	226/18.3/135
10	1,4,5,6,10, 11,36,43,115	1,2,3,4,5, 6,7,10,43	1,2,3,4,5,6, 7,8,9,10,146	305/24.7/130

number is below 7. However, we can see that MMoSU has missed 1/2/2 endmembers in the cases Num.=8/9/10, while PMoSU only misses one endmember among all the cases. This phenomenon is consistent with our analysis: when the noise is strong, MUSIC may ignore real endmembers. When the materials are highly mixed (endmember number is large), the situation will become even worse, because a single spectral signal will account for a smaller proportion of the total. Our previous work [42] has discussed that endmember missing will give rise to much more serious effects on the abundance inversion results than endmember redundancy. In other words, recall is more important than precision in unmixing problem. Therefore, we may conclude that the proposed simultaneous pruning strategy is superior to prepruning approaches under certain circumstances.

The second analysis is about weakly Pareto optimal. As is discussed in Section II-D, Tchebycheff-based decomposition may result in weakly Pareto optimal. Fig. 5 is an illustration, where the results on synthetic data 2 with five endmembers are taken for example. Each point corresponds to an individual on POF. Here, we only show 2-D figures with $f_1(\cdot)$ and $f_2(\cdot)$ for simplicity. It is observed that Fig. 5(a) has presented obvious weakly Pareto optimums. Theoretically, all the individuals in Fig. 5(a) are equally optimal, even though it seems that some individuals are completely below others. The reason is that the Tchebycheff decomposition will generate a rectangular contour on which all the points share the same Tchebycheff distance (rather than Euclidean distance) to the ideal point. It is worth noting that although it seems that we can directly use the bottom points on each column as the final solutions, this manner is unreasonable. If selecting the bottom points as the final solutions, the Euclidean distance is implicitly used. However, Tchebycheff distance is used during the whole optimization process. In this case, the final solutions of a formula calculated by Tchebycheff distance cannot be determined by minimizing the Euclidean distance.

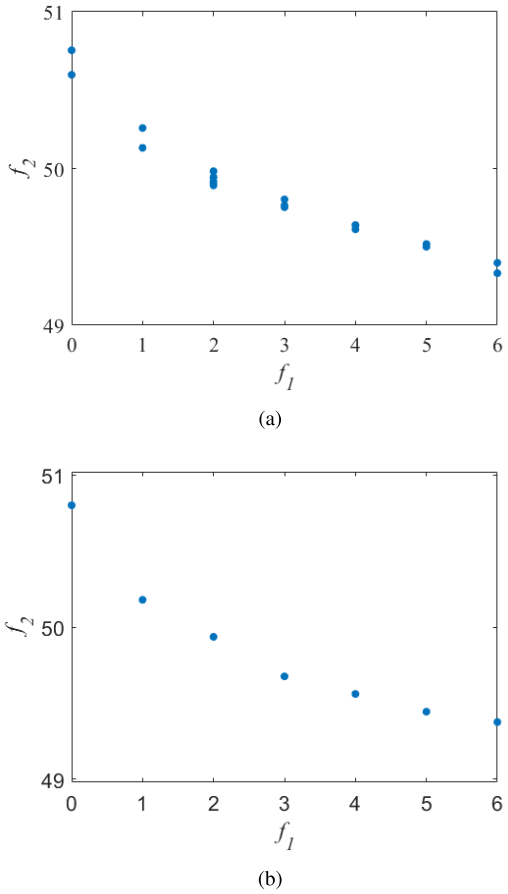


Fig. 5. Illustration about the weakly Pareto optimal. We take the results on synthetic data 2 with five endmembers and 25-dB noise for example. (a) PMoSU with Tchebycheff decomposition. (b) PMoSU with APBI decomposition. Each point corresponds to an individual on POF.

Minimizing the Euclidean distance makes no difference to random selection.

To overcome the weakly Pareto optimal problem, PMoSU proposes an improved APBI approach to replace the Tchebycheff decomposition, and the results are shown in Fig. 5(b). Obviously, the weakly Pareto optimums are eliminated. Similar results can also be observed in other data sets. Note that the solution obtained by APBI may not necessarily outperform every solution of Tchebycheff decomposition. Eliminating the weakly Pareto optimums is to guarantee the uniqueness of the solution, rather than improve the unmixing performance. We utilize APBI only to get the theoretically unique solution.

V. CONCLUSION

In this article, we propose a simultaneous sparse unmixing and library pruning method for hyperspectral imagery. The proposed method is motivated by the effectiveness of pruning operation. Research has demonstrated that pruning can reduce the mutual coherence of the spectral library and thus enhance the unmixing accuracy. However, traditional two-stage methods usually consider pruning as a preprocessing operation, which may lead to endmembers missing when the noise is strong. In PMoSU, we integrate the pruning and endmember extraction into a uniform process via transforming the pruning

to an objective function. To avoid the manually balance among different objectives, we use a multiobjective-based method for optimization. Furthermore, considering that the discrete range of sparse unmixing problem may lead to weakly Pareto optimal, we improve the optimization process by APBI. Experiments indicate that the proposed method has achieved our targets.

REFERENCES

[1] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

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

[3] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[4] 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.

[5] T. Chan, W. Ma, A. Ambikapathi, and C. Chi, "A simplex vol. maximization, framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.

[6] L. Drumetz, J. Chanussot, C. Jutten, W.-K. Ma, and A. Iwasaki, "Spectral variability aware blind hyperspectral image unmixing based on convex geometry," 2019, *arXiv:1904.03888*. [Online]. Available: http://arxiv.org/abs/1904.03888

[7] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

[8] Y. Altmann, M. Pereyra, and S. McLaughlin, "Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 174–185, Sep. 2015.

[9] Y. Woodbridge, U. Okun, G. Elidan, and A. Wiesel, "Unmixing K-Gaussians with application to hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7281–7293, Sep. 2019.

[10] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

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

[12] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[13] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

[14] X.-R. Feng, H.-C. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6245–6257, Oct. 2018.

[15] 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.

[16] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained $\ell_p - \ell_2$ optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142–1146, Sep. 2013.

[17] Y. E. Salehani, S. Gazor, and M. Cheriet, "Sparse hyperspectral unmixing via heuristic ℓ_p-norm approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1191–1202, Apr. 2018.

[18] C. Y. Zheng, H. Li, Q. Wang, and C. L. Philip Chen, "Reweighted sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 479–488, Jan. 2016.

[19] 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.

[20] 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.

[21] M.-D. Iordache, J. M. 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.

[22] R. Wang, H.-C. Li, A. Pizurica, J. Li, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using double reweighted sparse regression and total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1146–1150, Jul. 2017.

[23] S. Zhang, J. Li, H.-C. Li, C. Deng, and A. Plaza, "Spectral–spatial weighted sparse regression for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3265–3276, Jun. 2018.

[24] X. Xu, J. Li, C. Wu, and A. Plaza, "Regional clustering-based spatial preprocessing for hyperspectral unmixing," *Remote Sens. Environ.*, vol. 204, pp. 333–346, Jan. 2018.

[25] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, and C. Richard, "A fast multiscale spatial regularization for sparse hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 598–602, Apr. 2019.

[26] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "Learning a low-coherence dictionary to address spectral variability for hyperspectral unmixing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 235–239.

[27] A. Zare and K. C. 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.

[28] X. Xu, X. Tong, A. Plaza, Y. Zhong, H. Xie, and L. Zhang, "Joint sparse sub-pixel mapping model with endmember variability for remotely sensed imagery," *Remote Sens.*, vol. 9, no. 1, p. 15, Dec. 2016.

[29] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Dictionary pruning in sparse unmixing of hyperspectral data," in *Proc. 4th Workshop Hyperspectral Image Signal Process. (WHISPERS)*, Jun. 2012, pp. 1–4.

[30] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Potential and limitations of band selection and library pruning in sparse hyperspectral unmixing," in *Proc. 7th Workshop Hyperspectral Image Signal Process. Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.

[31] M.-D. Iordache, J. M. Bioucas-Dias, A. Plaza, and B. Somers, "MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4364–4382, Jul. 2014.

[32] M.-D. Iordache, L. Tits, J. M. Bioucas-Dias, A. Plaza, and B. Somers, "A dynamic unmixing framework for plant production system monitoring," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2016–2034, Jun. 2014.

[33] 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.

[34] H. Lin, X. Zhang, and W. Sun, "Spectral library pruning method in hyperspectral sparse unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 6561–6564.

[35] X. Zhang *et al.*, "Hyperspectral unmixing via low-rank representation with space consistency constraint and spectral library pruning," *Remote Sens.*, vol. 10, no. 2, p. 339, Feb. 2018.

[36] S. Das and A. Routray, "Covariance similarity approach for semiblind unmixing of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 937–941, Jun. 2019.

[37] 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.

[38] 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.

[39] R. Liu, B. Du, and L. Zhang, "Multiobjective optimized endmember extraction for hyperspectral image," *Remote Sens.*, vol. 9, no. 6, p. 558, Jun. 2017.

[40] 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.

[41] 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.

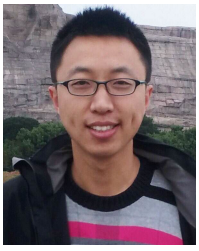
[42] X. Xu, Z. Shi, B. Pan, and X. Li, "A classification-based model for multi-objective hyperspectral sparse unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9612–9625, Dec. 2019.

[43] 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.

[44] X. Xu, L. Wang, B. Pan, and Z. Shi, "Robust sparse hyperspectral unmixing based on multi-objective optimization," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 5760–5763.
[45] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[46] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
[47] K. Deb, *Multiobjective Optimization Using Evolutionary Algorithms*. Hoboken, NJ, USA: Wiley, 2001.
[48] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
[49] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Dec. 2007.
[50] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
[51] J. Branke, K. Deb, H. Dierolf, and M. Osswald, "Finding knees in multi-objective optimization," in *Parallel Problem Solving From Nature-PPSN VIII*. Berlin, Germany: Springer, 2004, pp. 722–731.
[52] X. Zhang, Y. Tian, and Y. Jin, "A knee point-driven evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 19, no. 6, pp. 761–776, Dec. 2015.

Xia Xu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Astronautics, Beihang University, Beijing, China, in 2019.

She is an Assistant Professor with the College of Computer Science, Nankai University, Tianjin, China. Her research interests include hyperspectral unmixing, multiobjective optimization, and remote sensing image processing.



Bin Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively.

Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include machine learning, remote sensing image processing, and multiobjective optimization.

Zongqing Chen received the Ph.D. degree in mathematics from Nankai University, Tianjin, China, in 2016.

Since 2018, he has been a Post-Doctoral Researcher with the School of Statistics and Data Science, Nankai University. His research interests include graph theory and its applications, statistical inference, and machine learning theory.

Zhenwei Shi (Member, IEEE) 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 a Professor and the Dean of 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 the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the 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 the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) and the 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.

Tao Li received the Ph.D. degree in computer science from Nankai University, Tianjin, China, in 2007.

He is with the College of Computer Science, Nankai University, as a Professor. His main research interests include heterogeneous computing, machine learning, and Internet of Things.

Dr. Li is a Distinguished Member of the CCF.