Hyperspectral Unmixing Via Nonconvex Sparse and Low-Rank Constraint

Hongwei Han , Guxi Wang , Maozhi Wang , Jiaqing Miao , Si Guo, Ling Chen, Mingyue Zhang, and Ke Guo

Abstract—In recent years, sparse unmixing has attracted significant attention, as it can effectively avoid the bottleneck problems associated with the absence of pure pixels and the estimation of the number of endmembers in hyperspectral scenes. The joint-sparsity model has outperformed the single sparse unmixing method. However, the joint-sparsity model might cause some aliasing artifacts for the pixels on the boundaries of different constituent endmembers. To address this shortcoming, researchers have developed many unmixing algorithms based on low-rank representation, which makes good use of the global structure of data. In addition, the high mutual coherence of spectral libraries strongly affects the applicability of sparse unmixing. In this study, adopting combined constraints imposing sparsity and low rankness, a novel algorithm called nonconvex joint-sparsity and low-rank unmixing with dictionary pruning is developed In particular, we impose sparsity on the abundance matrix using the $\ell_{2,p}$ mixed norm, and we also employ the weighted Schatten p-norm instead of the convex nuclear norm as an approximation for the rank. The key parameter p is set between 0.4 and 0.6, and a good quality sparse solution is generated. The effectiveness of the proposed algorithm is demonstrated on both simulated and real hyperspectral datasets.

Index Terms—Hyperspectral images, joint-sparsity regression, low-rank representation (LRR), sparse unmixing, weighted Schatten p-norm.

Hongwei Han is with the Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu 610059, China, and also with the Engineering and Technical College of Chengdu University of Technology, Leshan 614000, China (e-mail: hhw_666666@163.com).

Guxi Wang is with the College of Architecture and Environment, Sichuan University, Chengdu 610000, China, and also with the National Institute of Measurement and Testing Technology, Chengdu 610000, China (e-mail: wanggx85@sina.com).

Maozhi Wang, Ling Chen, Mingyue Zhang, and Ke Guo are with the Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu 610059, China (e-mail: wangmz@cdut.edu.cn; 1493086926@qq.com; zhangmingyue@stu.cdut.edu.cn; guoke@cdut.edu.cn).

Jiaqing Miao is with the School of Computer Science and Technology, Southwest Minzu University, Chengdu 610000, China, and also with the Engineering and Technical College of Chengdu University of Technology, Leshan 614000, China (e-mail: mjq_011114117@163.com).

Si Guo is with the School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 610000, China, and also with the CCDC Geological Exploration and Development Research Institute, Chengdu 610000, China (e-mail: guosi123456@126.com).

Digital Object Identifier 10.1109/JSTARS.2020.3021520

I. Introduction

HYPERSPECTRAL imagery has high spectral resolution since it contains tens to hundreds of consecutive narrow spectral band images [1], [2]. In recent years, hyperspectral data have applications in all aspects of earth science and have become important constituents in the fields of applied biological analysis, pharmaceutical process monitoring and quality control [3], environmental monitoring, military exploration [4], quality control, and risk response [5]. However, mixed pixels typically exist in hyperspectral images due to their low spectral resolution and the complexity of material distribution over the earths surface, which seriously restrict the accurate interpretation of hyperspectral data [6], [7].

Spectral unmixing is an important technique for hyperspectral data exploitation [8], and it is divided into two parts: endmember extraction and abundance estimation. Endmember extraction infers the pure spectral signatures that constitute the mixed pixel, and abundance estimation calculates their corresponding fractions, called abundances [9]–[12]. The linear mixing model (LMM) is widely adopted for the spectral unmixing problem because of its computational tractability and flexibility [13]–[18].

In recent years, sparse unmixing has attracted much attention, as it can effectively avoid the problems associated with the absence of pure pixels and the estimation of the number of endmembers in hyperspectral scenes [19]. It searches for an optimal subset in a (potentially very large) spectral library, and each mixed pixel can be expressed as a linear combination of this optimal subset in these scenes [20].

Sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) was developed in [21]. It achieves higher accuracy and requires a shorter execution time than the previous methods. However, the sparsity-based unmixing algorithms cannot always obtain spectral unmixing results superior to those of traditional unmixing techniques [22]. There are two main reasons for this. On the one hand, the high mutual coherence of spectral libraries limits the performance of sparse mixing techniques [19]. On the other hand, many sparse unmixing methods only consider spectral information and ignore the rich spatial information between each pixel and its neighbors.

To mitigate these drawbacks, Iordache *et al.* proposed the sparse unmixing via variable splitting augmented Lagrangian and total variation (SUnSAL-TV) algorithm [8], which exploits the spatial contextual information by adding total variation regularization and spectral information simultaneously. The collaborative sparse unmixing via variable splitting and augmented Lagrangian (CLSUnSAL) algorithm was developed by Iordache

et al. [20]. This method improves the unmixing results by solving a jointly sparse regression problem, where the sparsity is simultaneously imposed on all pixels in the dataset, and it is faster than SUnSAL-TV in terms of running speed, since only one regularization parameter is used. However, these two algorithms neglect the influence of the high correlation of the spectral library for the unmixing results.

To mitigate the influence of the high correlation of the spectral library, Iordache *et al.* proposed the multiple signal classification (MUSIC) and collaborative sparse regression algorithm [19], which not only mitigates the coherence drawback, but also greatly reduces the running time of the CLSUnSAL algorithm. However, an actual spectral signature in the scene shows a mismatch with its corresponding dictionary sample in practice [23]. Fu *et al.* developed a dictionary-adjusted nonconvex sparsity-encouraging regression algorithm [23], which is effective in mitigating the undesirable spectral signature mismatch effects.

In recent years, low-rank representation (LRR) has been proved to make good use of the global structure of data and is currently widely used in sparse unmixing [24]–[27]. Qu *et al.* [28] sought abundance matrices utilizing the mixed $\ell_{2,1}$ norm and the LRR model to better capture the spatial data structure. Giampouras *et al.* [29] proposed the incremental proximal sparse low-rank unmixing algorithm and the alternating-direction sparse and low-rank unmixing (ADSpLRU) algorithm by imposing simultaneous sparsity and low rankness on the abundance matrix. Recently, Zhang *et al.* [22] have developed an unmixing algorithm via LRR based on the space consistency constraint with spectral library pruning. In [10], a new algorithm called joint-sparse blocks and low-rank unmixing was proposed via concatenating the proposed joint-sparsity block structure and low-rankness assumption on the abundance coefficients. Considering the local spatial correlation of hyperspectral images, a new unmixing method with superpixel segmentation and LRR based on the Gaussian mixture model was proposed [30]. Li *et al.* [31] proposed a novel sparse unmixing algorithm named superpixel-based reweighted low-rank and total variation. By incorporating a low-rank regularization to the traditional sparse unmixing models, Xu [32] developed a generic unmixing model. To utilize nonlocal self-similarity property, Zheng *et al.* [33] propose a sparse unmixing model with nonlocal low-rank prior. In [34], a weighted nonlocal low-rank tensor decomposition method for sparse unmixing was proposed to use spatial and spectral information fully. The unmixing models based on low rank have achieved better performance than those with only sparse constraints.

A previous work [28] proposed a low-rank model of abundance estimation, dictionary of which used in unmixing satisfies the full column rank property. In order to make use of this property, the dictionary was constructed from the pure endmembers extracted by the vertex component analysis (VCA) algorithm [35]. However, VCA should satisfy the assumption of the existence of pure pixels in datasets, which is generally not true in practice. Recently, the spectral library was used for sparse unmixing based on LRR instead of the endmember dictionary to avoid extracting the endmembers from the original hyperspectral data [10]. However, the high mutual coherence of the library signatures causes this model to perform poorly [22].
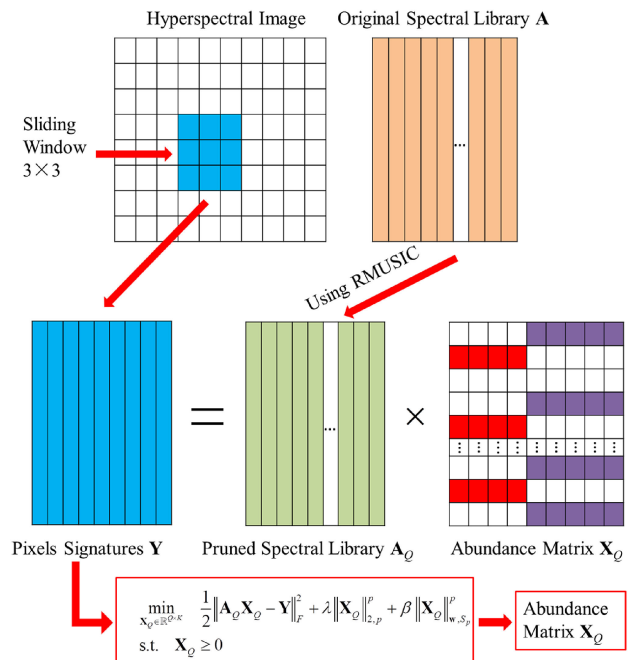

Fig. 1. Flowchart of the proposed method. The abundance matrix $\mathbf{X}_Q$ is considered sparse and of low rank (here, rank $= 2$). The white cells in matrix $\mathbf{X}_Q$ represent zero values.

To mitigate this drawback, we propose a low-rank model of abundance estimation based on the dictionary pruning strategy.

In a hyperspectral image, the joint-sparsity constraint [36] provides a more general assumption that neighboring pixels are composed of similar endmembers but do not necessarily have similar abundance for the same endmember [20]. However, the joint-sparsity model might cause some aliasing artifacts for the pixels on the boundaries of different constituent endmembers [28]. Further investigation shows that high spatial similarity is associated with the low column rank attribute of data. That is, not only does the abundance matrix have a low rank, but it also has low sparsity [28]. Thus, inspired by [29], we propose a novel spectral unmixing model, which simultaneously imposes sparsity and low rankness on abundance matrices for pixels in a sliding window, as in [29]. First, we prune the spectral library with the robust MUSIC (RMUSIC) strategy for dictionary pruning in [23], which can be implemented quite efficiently. After this step, we can obtain a pruned spectral library, the size of which is usually much smaller than that of the original. Next, the pruned spectral library can be used in the proposed algorithm. Second, we improve the jointly sparse regression problem using the mixed $\ell_{2,p}(0 < p < 1)$ norm, and we also employ the weighted Schatten p-norm [37] as an approximation for the rank instead of the convex nuclear norm [10], [29]. Thus, a novel model called nonconvex sparse and low-rank unmixing with dictionary pruning (NCJSpLRUDP) is developed. Moreover, the proposed algorithm follows the framework of the classic alternating direction method of multipliers (ADMM) [11], [19], [20], which can turn a complex problem into a series of simple problems; it has been successfully used in [8], [10], and [29] for global convergence guarantees. The results of the simulated and real-data experiments indicate the efficacy of the proposed algorithm. Fig. 1 presents a summary of the entire process.

The remainder of this article is organized as follows. Section II reviews the jointly sparse regression and some related methods. Section III describes the proposed method and derives the proposed algorithm. Section IV analyzes the performance of the proposed approach with simulated data and real hyperspectral data and discusses the parameter selection. Section V concludes this article with some important remarks.

II. BASIC METHODS CONTRIBUTING TO THE PROPOSED APPROACH

In this section, we first outline the jointly sparse regression. Then, two basic methods for our algorithm, abundance estimation via LRR [28] and the RMUSIC strategy for dictionary pruning [23], are introduced. Finally, we discuss the weighted Schatten p-norm minimization (WSNM) model [37].

A. Jointly Sparse Regression

LMM is widely adopted for the spectral unmixing problem [13]. Jointly sparse regression based on LMM has been well studied because of its good unmixing performance [10], [20], [29]. We review this method in the following.

Let $\mathbf{Y} \in \mathbb{R}^{L \times n}$ be the observed dataset containing n pixels with L spectral bands and $\mathbf{X} \in \mathbb{R}^{m \times n}$ is the abundance fraction matrix. Let $\mathbf{A} \in \mathbb{R}^{L \times m}$ be a spectral library composed of m spectral signatures. The model can be described as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{1}$$

where $\mathbf{N} \in \mathbb{R}^{L \times n}$ is the noise and model error. In the hyperspectral community, the constraints $\mathbf{X} \geq \mathbf{0}$ and $\mathbf{1}^T\mathbf{X} = \mathbf{1}^T$ are often imposed on $\mathbf{X}$ [10]. We term $\mathbf{X} \geq \mathbf{0}$ as the abundance nonnegativity constraint and $\mathbf{1}^T\mathbf{X} = \mathbf{1}^T$ as abundance sum-to-one constraint [16]. By introducing a regularization parameter, the jointly sparse unmixing model can be written as

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{\text{row},0} \quad \text{s.t.} \quad \mathbf{X} \geq \mathbf{0} \tag{2}$$

where $\|\mathbf{X}\|_{\text{row},0}$ denotes the number of nonzero rows in $\mathbf{X}$. However, the optimization problem in (2) is NP-hard. In general, $\|\mathbf{X}\|_{\text{row},0}$ is convexly relaxed to $\|\mathbf{X}\|_{2,1} = \sum_{k=1}^{m} \|\mathbf{x}^k\|_2$ (where $\mathbf{x}^k$ denotes the kth line of $\mathbf{X}$). Then, (2) can be expressed as

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,1} \quad \text{s.t.} \quad \mathbf{X} \geq \mathbf{0} \tag{3}$$

where $\|\mathbf{X}\|_{2,1}$ denotes the $\ell_{2,1}$ mixed norm, which promotes sparsity among the lines of $\mathbf{X}$. This is because $\|\mathbf{X}\|_{2,1}$ norm imposes sparsity among the endmembers simultaneously for all pixels [20].

B. RMUSIC Strategy for Dictionary Pruning

As discussed in Section I, high mutual coherence of the dictionary and the large dictionary size strongly limit the performance of sparse unmixing. To alleviate this situation, dictionary pruning can be applied. RMUSIC is an effective dictionary pruning strategy [23]. To introduce this method, we start with the subspace-based dictionary pruning method called MUSIC [19]. We consider the subspace method in the noiseless case $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Let $\mathbf{U}_S \in \mathbb{R}^{L \times t}$ denote a matrix, which contains the

first t left singular vectors of $\mathbf{Y}$. Under some mild assumptions (see [19] for details), we have

$$P_{\mathbf{U}_S}^{\perp}\mathbf{a}_j = 0 \text{ if } \mathbf{a}_j = \mathbf{a}_{j_T} \text{ for some } T \in \{1, \ldots, t\} \tag{4}$$

where $P_{\mathbf{U}_S}^{\perp} = \mathbf{I} - \mathbf{U}_S\mathbf{U}_S^T$ is the projector on range$(\mathbf{U}_S)^{\perp}$. The physical meaning of (4) is that if a spectral sample $\mathbf{a}_j$ in the dictionary is also one of the spectral signatures in the scene, then it must be perpendicular to the orthogonal complement signal subspace [23]. However, noise does exist in hyperspectral data. Then, the left-hand side of (4) cannot always be true. We can estimate the endmembers with the following improved formulation:

$$\Upsilon_{\text{MUSIC}}(j) = \frac{\mathbf{a}_j^T P_{\mathbf{U}_S}^{\perp}\mathbf{a}_j}{\|\mathbf{a}_j\|_2^2}. \tag{5}$$

We determine $\hat{\Lambda} = \{\hat{j}_1, \ldots, \hat{j}_t\}$ for $T = 1, \ldots, t$ and then $\Upsilon_{\text{MUSIC}}(\hat{j}_T) < \Upsilon_{\text{MUSIC}}(j)$ for all $j \notin \hat{\Lambda}$.

Now, we consider the presence of dictionary mismatches and propose to replace (5) by the following RMUSIC metric:

$$\Upsilon_{\text{MUSIC}}(j) = \frac{(\mathbf{a}_j - \xi)^T P_{\mathbf{U}_S}^{\perp}(\mathbf{a}_j - \xi)}{\|(\mathbf{a}_j - \xi)\|_2^2}$$

$$\text{s.t.} \quad \|\xi\|_2 \leq \delta \tag{6}$$

where $\delta > 0$ is prespecified. It adjusts the dictionary to find a better match in the subspace scene.

C. Abundance Estimation via LRR

A previous study [28] reported that the joint-sparsity model introduces some aliasing artifacts on the boundaries when the neighboring pixels consist of different endmembers. To mitigate this drawback, the LRR model for abundance estimation was proposed by the following theorem [38].

Theorem 1: Assume matrices $\mathbf{Y} \in \mathbb{R}^{L \times n}$, $\mathbf{A} \in \mathbb{R}^{L \times m}$, and $\mathbf{X} \in \mathbb{R}^{m \times n}$, which satisfy $\mathbf{Y} = \mathbf{A}\mathbf{X}$. If rank$(\mathbf{Y}) = r \leq \min(m, n)$ and rank$(\mathbf{A}) = m$, then we have

$$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{Y}) = r. \tag{7}$$

In [28], dictionary $\mathbf{A}$ is made up of pure endmembers that are extracted from the hyperspectral image, and it usually satisfies the full column rank property. This is because these pure endmembers differ from each other, and the dimension of real image data L is much larger than the number of endmembers m. It is evident that matrix $\mathbf{Y}$ is a low-rank matrix if the columns of $\mathbf{Y}$ are highly correlated. Then, we can confirm that the corresponding abundance matrix $\mathbf{X}$ is also a low-rank matrix.

To use this property, we introduce the LRR method for solving the joint abundance estimation problem. Under a simple LMM, it is equivalent to solving the following optimization problem:

$$\mathbf{X}^* = \arg\min \text{rank}(\mathbf{X})$$

$$\text{s.t.} \quad \mathbf{Y} - \mathbf{A}\mathbf{X} = \mathbf{0}, \ \mathbf{X} \geq \mathbf{0}, \ \mathbf{1}^T\mathbf{X} = \mathbf{1}^T \tag{8}$$

where $\mathbf{X}^*$ is the lowest rank solution, and $\mathbf{Y} \in \mathbb{R}^{L \times n}$ refers to the given data.

Algorithm 1: WSNM Via GST.

1. **Input:** $\mathbf{Y}$, ω_i in ascending order, p
2. $\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^T$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots \sigma_r)$
3. **For** $i = 1$ to r **do**
4. $\delta_i = \mathrm{GST}(\sigma_i, \omega_i, p)$
5. **End**
6. $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_r)$
7. **Return** $\hat{\mathbf{X}} = \mathbf{U}\Delta\mathbf{V}^T$
8. **Output:** $\mathbf{X}$

D. Weighted Schatten p-Norm Minimization

In a previous work [37], the authors proposed the WSNM model, which provides a better approximation to the original low-rank assumption than nuclear norm minimization. We introduce this model briefly here.

The weighted Schatten p-norm of matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ is defined as

$$\|\mathbf{X}\|_{\mathbf{w}, S_p} = \left(\sum_{i=1}^{\min\{n,m\}} \omega_i \sigma_i^p \right)^{\frac{1}{p}} \tag{9}$$

where $\mathbf{w} = [\omega_1, \ldots, \omega_{\min\{n,m\}}]$ is a nonnegative vector, and σ_i is the ith singular value of $\mathbf{X}$. Given a matrix $\mathbf{Y}$, the WSNM problem is described as follows:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X}\|_{\mathbf{w}, S_p}^p \tag{10}$$

where λ is a tradeoff parameter. The above WSNM can be equivalently transformed into independent nonconvex ℓ_p norm subproblems, whose global optimum can be efficiently solved by the generalized soft-thresholding (GST) algorithm [39]. Then, the WSNM algorithm is summarized in Algorithm 1, where GST solves the ℓ_p norm minimization.

Let $f(x) = (1\,2)(x - y)^2 + \lambda|x|^p$. Given p and λ, the thresholding value τ_p^{GST} is

$$\tau_p^{\mathrm{GST}}(\lambda) = (2\lambda(1 - p))^{\frac{1}{2-p}} + \lambda p(2\lambda(1 - p))^{\frac{p-1}{2-p}}. \tag{11}$$

Then, the GST functions as follows:

$$T_p^{\mathrm{GST}}(y; \lambda) = \begin{cases} 0, & \text{if } |y| \leq \tau_p^{\mathrm{GST}}(\lambda) \\ \mathrm{sgn}(y) S_p^{\mathrm{GST}}(|y|; \lambda), & \text{if } |y| > \tau_p^{\mathrm{GST}}(\lambda) \end{cases} \tag{12}$$

where $S_p^{\mathrm{GST}}(y; \lambda)$ is the unique minimum of $f(x)$ in the range of $(\tau_p^{\mathrm{GST}}(\lambda), +\infty)$. For more details about the GST algorithm, see [39].

III. PROPOSED METHOD: NONCONVEX SPARSE AND LOW-RANK UNMIXING VIA DICTIONARY PRUNING

In this section, we first propose the nonconvex sparse and low-rank models based on the jointly sparse regression, and we then design our unmixing algorithm under the ADMM framework.

A. Problem Formulation

In order to make better use of the low rank, we also consider a small sliding square window that contains K adjacent pixels ($K = l \times l$) following [28] and [29]. Let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_K]$ be the $L \times K$ matrix, where L is the number of bands and K is the number of pixels in the window. Let $\mathbf{A} \in \mathbb{R}^{L \times m}$ be a spectral library containing m spectral signatures. Under LMM, the mixing process can be modeled as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{13}$$

where $\mathbf{X} \in \mathbb{R}^{m \times K}$ is the fractional abundance matrix, and $\mathbf{N} \in \mathbb{R}^{L \times K}$ is a noise matrix. From Section II-A, the $\ell_{2,1}$ matrix norm is often used to find the jointly sparse solution. It is worth noting that the $\ell_{2,1}$ norm is a convex relaxation of a row-ℓ_0 quasi-norm [10]. In fact, computational studies have shown that the $\ell_p (0 < p < 1)$ norm is sparser than the ℓ_1 norm [1], [2], [40]–[43]. The generalized $\ell_{2,p}$ norm [10], [23] naturally has better sparsity than $\ell_{2,1}$. In this study, we impose sparsity on the abundance matrix $\mathbf{X}$ using the $\ell_{2,p}$ mixed norm. Then, we obtain the following proposed collaborative sparse regression model:

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times K}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,p}^p \quad \text{s.t.} \quad \mathbf{X} \geq \mathbf{0} \tag{14}$$

where λ is the regularization parameter, and $\|\mathbf{X}\|_{2,p} = \left(\sum_1^m \|\mathbf{x}^i\|_2^p \right)^{\frac{1}{p}}$, $p \in (0, 1]$ (where $\mathbf{x}^i$ denotes the ith line of $\mathbf{X}$) is the $\ell_{2,p}$ mixed norm.

As described in Section II-C, the joint-sparsity model introduces some aliasing artifacts on the boundaries when the neighboring pixels consist of different endmembers. To mitigate this drawback, Qu *et al.* proposed the LRR model for abundance estimation, as it can better capture the spatial data structure by seeking the lowest rank representation [28]. Therefore, we add a low-rankness regularization term in (8) to the collaborative sparse regression model. Then, our optimization problem for spectral unmixing becomes as follows:

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times K}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,p}^p + \beta \, \mathrm{rank}\,(\mathbf{X})$$
$$\text{s.t.} \quad \mathbf{X} \geq \mathbf{0} \tag{15}$$

where $\lambda > 0$ and $\beta > 0$ are regularization parameters. In fact, (15) simultaneously imposes sparsity and low rankness on the abundance matrix for the pixels in a slide window. It is obvious that the optimization problem in (15) is nonconvex and difficult to solve.

To solve the above problem, the authors of [10] and [29] first proposed a convex relaxation by replacing the rank of $\mathbf{X}$ with the well-known nuclear norm $\|\mathbf{X}\|_*$ (the sum of the singular values). However, the nuclear norm regularization-based model shrinks too many of the low-rank components of the data [37]. Therefore, the weighted Schatten p-norm [37] is used as a surrogate for the rank instead of the convex nuclear norm. Extensive experimental results show that the WSNM problem provides a better approximation to the original low-rank matrix approximation problem.

Then, the surrogate optimization problem is written as

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times K}} \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_{2,p}^p + \beta \|\mathbf{X}\|_{\mathbf{w},S_p}^p$$

$$\text{s.t.} \quad \mathbf{X} \geq \mathbf{0} \tag{16}$$

where $\|\mathbf{X}\|_{\mathbf{w},S_p}$ is defined by (9). It is worth noting that $\mathbf{A}$ is a full-rank matrix that can guarantee the validity of the low-rank model above (discussed in Section II-C). Unfortunately, this condition is generally not true because the spectral library $\mathbf{A}$ is potentially very large. Therefore, we first prune the spectral library $\mathbf{A}$. Assume that $\mathbf{A}_Q \in \mathbb{R}^{L \times Q}(Q \leq m)$ is the pruned dictionary, which is obtained by the RMUSIC method described in Section II-B, and $\mathbf{X}_Q \in \mathbb{R}^{Q \times K}$ is the corresponding abundance fraction matrix. Then, (16) can be rewritten as

$$\min_{\mathbf{X}_Q} \frac{1}{2} \|\mathbf{A}_Q\mathbf{X}_Q - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}_Q\|_{2,p}^p + \beta \|\mathbf{X}_Q\|_{\mathbf{w},S_p}^p$$

$$\text{s.t.} \quad \mathbf{X}_Q \geq \mathbf{0}. \tag{17}$$

Although the above model is nonconvex, it can be effectively solved with the ADMM. Next, we solve the proposed model in (17) using the ADMM.

B. Nonconvex Sparse and Low-Rank Unmixing Via Pruning Dictionary Algorithm

Instead of directly solving the model in (17), we use the regularization method to transform the above equation into an unconstrained optimization problem, and then, we use the ADMM [8], [10], [11] to solve it. After embedding the nonnegativity in (17), we obtain the following equivalent form:

$$\min_{\mathbf{X}_Q} \frac{1}{2} \|\mathbf{A}_Q\mathbf{X}_Q - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}_Q\|_{2,p}^p$$

$$+ \beta \|\mathbf{X}_Q\|_{\mathbf{w},S_p}^p + l_{R+}(\mathbf{X}_Q) \tag{18}$$

where $l_{R+}(\mathbf{X}_Q)$ is the indicator function, which is zero if $l_{R+}(\mathbf{X}_Q)$ is nonnegative and $+\infty$ otherwise. We introduce four variables ($\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3,$ and $\mathbf{V}_4$) and obtain

$$\min_{\mathbf{X}_Q} \frac{1}{2} \|\mathbf{A}_Q\mathbf{X}_Q - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}_Q\|_{2,p}^p$$

$$+ \beta \|\mathbf{X}_Q\|_{\mathbf{w},S_p}^p + l_{R+}(\mathbf{X}_Q)$$

$$\text{s.t.} \ \mathbf{V}_1 = \mathbf{A}_Q\mathbf{X}_Q, \mathbf{V}_2 = \mathbf{X}_Q, \mathbf{V}_3 = \mathbf{X}_Q, \mathbf{V}_4 = \mathbf{X}_Q \tag{19}$$

We let

$$g(\mathbf{V}) = \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{V}_2\|_{2,p}^p$$

$$+ \beta \|\mathbf{V}_3\|_{\mathbf{w},S_p}^p + l_{R+}(\mathbf{V}_4) \tag{20}$$

where $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4]^T$, and we define

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_Q \\ \mathbf{I} \\ \mathbf{I} \\ \mathbf{I} \end{bmatrix}, B = \begin{bmatrix} -\mathbf{I} & 0 & 0 & 0 \\ 0 & -\mathbf{I} & 0 & 0 \\ 0 & 0 & -\mathbf{I} & 0 \\ 0 & 0 & 0 & -\mathbf{I} \end{bmatrix}. \tag{21}$$

Therefore, we obtain a compact form of (19) as

$$\min_{\mathbf{X}_Q,\mathbf{V}} g(\mathbf{V}) \quad \text{s.t.} \quad \mathbf{G}\mathbf{X}_Q + \mathbf{B}\mathbf{V} = \mathbf{0}. \tag{22}$$

By introducing augmented Lagrangian multipliers $\mathbf{D} = [\mathbf{D}_1^T, \mathbf{D}_2^T, \mathbf{D}_3^T, \mathbf{D}_4^T]^T$, the optimization problem (18) is converted to solve the following Lagrangian function minimization problem:

$$\mathcal{L}(\mathbf{X}_Q, \mathbf{V}, \mathbf{D}) = g(\mathbf{V}) + \frac{\mu}{2} \|\mathbf{G}\mathbf{X}_Q + \mathbf{B}\mathbf{V} - \mathbf{D}\|_F^2 \tag{23}$$

where μ is a positive penalty parameter. Then, the steps of the ADMM are derived as

$$\begin{cases} \mathbf{X}_Q^{(k+1)} = \arg\min_{\mathbf{X}_Q} \mathcal{L}(\mathbf{X}_Q, \mathbf{V}^{(k)}, \mathbf{D}^{(k)}) \\ \mathbf{V}^{(k+1)} = \arg\min_{\mathbf{V}} \mathcal{L}(\mathbf{X}_Q^{(k+1)}, \mathbf{V}, \mathbf{D}^{(k)}) \\ \mathbf{D}^{(k+1)} = \mathbf{D}^{(k)} - \mathbf{G}\mathbf{X}_Q^{(k+1)} - \mathbf{B}\mathbf{V}^{(k+1)} \end{cases}. \tag{24}$$

Now, we detail each subproblem of (24) as follows: The subproblem with respect to $\mathbf{X}_Q$ gives

$$\mathbf{X}_Q^{(k+1)} = \arg\min_{\mathbf{X}_Q} \mathcal{L}(\mathbf{X}_Q, \mathbf{V}^{(k)}, \mathbf{D}^{(k)})$$

$$= \arg\min_{\mathbf{X}_Q} \frac{\mu}{2} \left\|\mathbf{A}_Q\mathbf{X}_Q - \mathbf{V}_1^{(k)} - \mathbf{D}_1^{(k)}\right\|_F^2$$

$$+ \frac{\mu}{2} \left\|\mathbf{X}_Q - \mathbf{V}_2^{(k)} - \mathbf{D}_2^{(k)}\right\|_F^2$$

$$+ \frac{\mu}{2} \left\|\mathbf{X}_Q - \mathbf{V}_3^{(k)} - \mathbf{D}_3^{(k)}\right\|_F^2$$

$$+ \frac{\mu}{2} \left\|\mathbf{X}_Q - \mathbf{V}_4^{(k)} - \mathbf{D}_4^{(k)}\right\|_F^2. \tag{25}$$

The solution of (25) is simply

$$\mathbf{X}_Q^{(k+1)} = \left(\mathbf{A}_Q^T\mathbf{A}_Q + 3\mathbf{I}\right)^{-1} \left(\mathbf{A}_Q^T \left(\mathbf{V}_1^{(k)} + \mathbf{D}_1^{(k)}\right)\right.$$

$$\left. + \mathbf{V}_2^{(k)} + \mathbf{D}_2^{(k)} + \mathbf{V}_3^{(k)} + \mathbf{D}_3^{(k)} + \mathbf{V}_4^{(k)} + \mathbf{D}_4^{(k)}\right). \tag{26}$$

The subproblem with respect to $\mathbf{V}_1$ is equivalent to the following problem:

$$\mathbf{V}_1^{(k+1)} = \arg\min_{\mathbf{V}_1} \frac{1}{2} \|\mathbf{V}_1 - \mathbf{Y}\|_F^2$$

$$+ \frac{\mu}{2} \left\|\mathbf{A}_Q\mathbf{X}_Q^{(k+1)} - \mathbf{V}_1 - \mathbf{D}_1^{(k)}\right\|_F^2. \tag{27}$$

Note that (27) has a closed-form solution

$$\mathbf{V}_1^{(k+1)} = \frac{1}{1 + \mu} \left[\mathbf{Y} + \mu \left(\mathbf{A}_Q\mathbf{X}_Q^{(k+1)} - \mathbf{D}_1^{(k)}\right)\right]. \tag{28}$$

For the $\mathbf{V}_2$ subproblem, we have

$$\mathbf{V}_2^{(k+1)} = \arg\min_{\mathbf{V}_2} \lambda\|\mathbf{V}_2\|_{2,p}^p$$

$$+ \frac{\mu}{2} \left\|\mathbf{X}_Q^{(k+1)} - \mathbf{V}_2 - \mathbf{D}_2^{(k)}\right\|_F^2. \tag{29}$$

Note that (29) can be solved by adapting the factored gradient approach and has been proved to converge toward a local

minimum [44]. The partial derivative of $\|\mathbf{V}_2\|_{2,p}^p$ with respect to element $\mathbf{V}_2[i,j]$ is

$$\frac{\partial \|\mathbf{V}_2\|_{2,p}^p}{\partial \mathbf{V}_2[i,j]} = p \|\mathbf{v}_2^i\|_2^{p-2} \mathbf{V}_2[i,j] \tag{30}$$

and the gradient of the objective function with respect to $\mathbf{V}_2$ is

$$\lambda \mathbf{P} \mathbf{V}_2 - \mu \left(\mathbf{X}_Q^{(k+1)} - \mathbf{V}_2 - \mathbf{D}_2^{(k)} \right) \tag{31}$$

where $\mathbf{P} = \operatorname{diag}(p\|\mathbf{v}_2^i\|_2^{p-2})$, and $\mathbf{v}_2^i$ is the ith row of $\mathbf{V}_2$. Then, we define $\mathbf{W}$ as $\mathbf{W} = \operatorname{diag}(p^{-1/2}\|\mathbf{v}_2^i\|_2^{1-p/2})$, and $\mathbf{W}^{-2} = \mathbf{P}$. After having replaced $\mathbf{W}^{-2} = \mathbf{P}$ in (31), the optimum solution of (29) is

$$\mathbf{V}_2^{(k+1)} = \mu \mathbf{W} \left(\mu \mathbf{W}^2 + \lambda \mathbf{I} \right)^{-1} \mathbf{W}^T \left(\mathbf{X}_Q^{(k+1)} - \mathbf{D}_2^{(k)} \right). \tag{32}$$

Next, $\mathbf{V}_3$ is computed by a shrinkage operation, that is,

$$\mathbf{V}_3^{(k+1)} = \arg\min_{\mathbf{V}_3} \beta \|\mathbf{V}_3\|_{\mathbf{w},S_p}^p$$
$$+ \frac{\mu}{2} \left\| \mathbf{X}_Q^{k+1} - \mathbf{V}_3 - \mathbf{D}_3^{(k)} \right\|_F^2. \tag{33}$$

This is the same as the WSNM problem in Section II-D. Then, we have

$$\mathbf{V}_3^{(k+1)} = \text{WSNM} \left(\mathbf{X}_Q^{(k+1)} - \mathbf{D}_3^{(k)}, \frac{2\beta}{\mu}, p \right). \tag{34}$$

The subproblem with respect to $\mathbf{V}_4$ gives

$$\mathbf{V}_4^{(k+1)} = \arg\min_{\mathbf{V}_4} l_{R_+} (\mathbf{V}_4)$$
$$+ \frac{\mu}{2} \left\| \mathbf{X}_Q^{(k+1)} - \mathbf{V}_4 - \mathbf{D}_4^{(k)} \right\|_F^2. \tag{35}$$

In (35), the role of the $l_{R_+}(\mathbf{X})$ term is to project the solution onto the nonnegative orthant, and the value of $\mathbf{V}_4$ is given by

$$\mathbf{V}_4^{(k+1)} = \max \left(\mathbf{X}_Q^{(k+1)} - \mathbf{D}_4^{(k)}, 0 \right). \tag{36}$$

Finally, the multipliers $\mathbf{D}$ are sequentially updated as follows:

$$\begin{cases} \mathbf{D}_1^{(k+1)} = \mathbf{D}_1^{(k)} - \mathbf{A}_Q \mathbf{X}_Q^{(k+1)} + \mathbf{V}_1^{(k+1)} \\ \mathbf{D}_i^{(k+1)} = \mathbf{D}_i^{(k)} - \mathbf{X}_Q^{(k+1)} + \mathbf{V}_i^{(k+1)}, \ i = 2, 3, 4 \end{cases}. \tag{37}$$

Under the ADMM framework, NCJSpLRUDP is summarized in Algorithm 2.

C. Initial and Convergence Conditions

The proposed method is obviously nonconvex. The initial condition must be chosen; otherwise, the algorithm can easily fall into the local minimum. In general, the setting of the initial values can be postulated by utilizing domain-specific knowledge in engineering applications. If no good starting points are available, then the minimum Frobenius norm solution, which was shown to be successful in [44], is a good initializer. Therefore, the minimum Frobenius norm solution is used for initialization in the proposed algorithm.

In the proposed algorithm, the convergence conditions are defined based on the primal and the dual ADMM variables [10],

---

Algorithm 2: NCJSpLRUDP.

1. **Input:** $\mathbf{Y}$ and original spectral $\mathbf{A}$
2. **Initialization:** $k = 0$, $0 < p \leq 1$, $\lambda \geq 0$, $\mu \geq 0$, $\mathbf{X}^{(0)}$, $\mathbf{V}_1^{(0)}, \mathbf{V}_2^{(0)}, \mathbf{V}_3^{(0)}, \mathbf{V}_4^{(0)}, \mathbf{D}_1^0, \mathbf{D}_2^0, \mathbf{D}_3^0, \mathbf{D}_4^0$
3. Estimate the pruned dictionary $\mathbf{A}_Q$ using the RMUSIC method.
4. **Repeat:**
5. $\mathbf{X}_Q^{(k+1)} = (\mathbf{A}_Q^T \mathbf{A}_Q + 3\mathbf{I})^{-1} \left(\mathbf{A}_Q^T \left(\mathbf{V}_1^{(k)} + \mathbf{D}_1^{(k)} \right) \right.$
   $\left. + \mathbf{V}_2^{(k)} + \mathbf{D}_2^{(k)} + \mathbf{V}_3^{(k)} + \mathbf{D}_3^{(k)} + \mathbf{V}_4^{(k)} + \mathbf{D}_4^{(k)} \right)$
6. $\mathbf{V}_1^{(k+1)} = \frac{1}{1+\mu} \left[\mathbf{Y} + \mu \left(\mathbf{A}_Q \mathbf{X}_Q^{(k+1)} - \mathbf{D}_1^{(k)} \right) \right]$
7. $\mathbf{V}_2^{(k+1)} = \mu \mathbf{W} \left(\mu \mathbf{W}^2 + \lambda \mathbf{I} \right)^{-1} \mathbf{W}^T \left(\mathbf{X}_Q^{(k+1)} - \mathbf{D}_2^{(k)} \right)$
8. $\mathbf{V}_3^{(k+1)} = \text{WSNM} \left(\mathbf{X}_Q^{(k+1)} - \mathbf{D}_3^{(k)}, \frac{2\beta}{\mu}, p \right)$
9. $\mathbf{V}_4^{(k+1)} = \max(\mathbf{X}_Q^{(k+1)} - \mathbf{D}_4^{(k)}, 0)$
10. **Update Lagrangian multipliers:**
    $\mathbf{D}_1^{(k+1)} = \mathbf{D}_1^{(k)} - \mathbf{A}_Q \mathbf{X}_Q^{(k+1)} + \mathbf{V}_1^{(k+1)}$
    $\mathbf{D}_i^{(k+1)} = \mathbf{D}_i^{(k)} - \mathbf{X}_Q^{(k+1)} + \mathbf{V}_i^{(k+1)}, \ i = 2, 3, 4$
11. **Update iteration:** $k = k + 1$
12. **Until** the stopping criterion is satisfied.
13. **Output:** Abundance matrix $\mathbf{X}$.

---

[20], [29]. Then, we define the primal $\mathbf{r}^{(k)}$ and the dual $\mathbf{d}^{(k)}$ as follows:

$$\mathbf{r}^{(k)} = \mathbf{G} \mathbf{X}_Q^{(k)} + \mathbf{B} \mathbf{V}^{(k)}$$

$$\mathbf{d}^{(k)} = \mu \mathbf{G}^T \mathbf{B} \left(\mathbf{V}^{(k)} - \mathbf{V}^{(k-1)} \right).$$

The proposed algorithm stops when the termination criterion

$$\left\| \mathbf{r}^{(k)} \right\|_2 \leq \zeta, \ \left\| \mathbf{d}^{(k)} \right\|_2 \leq \zeta$$

holds or when the maximum number of iterations is reached. In our experiments, we set $\zeta = \sqrt{(3m + L)K} \zeta^{\text{rel}}$, where the relative tolerance ζ^{rel} has been empirically determined as 10^{-4} [29]. Furthermore, the maximum number of iterations is set to 1000.

IV. EXPERIMENTS WITH SIMULATED DATA AND REAL HYPERSPECTRAL DATA

To validate the advantages of our method, we demonstrate the proposed algorithm in both the simulated and the real scenarios. We compare the proposed algorithm with four state-of-the-art unmixing algorithms, namely, SUnSAL [21], CLSUnSAL [20], SUnSAL-TV [8], and ADSpLRU [29].

For simulated test problems, three spectral libraries, $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$, are used to generate different synthetic hyperspectral images. $\mathbf{A}_1$ is Chapter 1 of the U.S. Geological Survey (USGS) spectral library (splib06a) [8], which comprises 498 spectral signatures with 224 spectral bands uniformly distributed in the

interval of 0.4–2.5 μm. $\mathbf{A}_2$ and $\mathbf{A}_3$ are randomly extracted from $\mathbf{A}_1$. $\mathbf{A}_2$ has 240 spectral signatures and $\mathbf{A}_3$ has 332 spectral signatures.

The unmixing performance is measured in terms of abundance estimation using the signal-to-reconstruction error (SRE) [8]

$$\text{SRE} = 10\log_{10}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\|\hat{\mathbf{x}}_i\|_2^2}{\frac{1}{n}\sum_{i=1}^{n}\|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2}\right)$$

where n is the number of pixels, and $\hat{\mathbf{x}}_i$ and $\mathbf{x}_i$ are the estimated and original abundance vectors of the ith pixel, respectively. We also use an evaluation metric called the root-mean-square error (RMSE), which is formally defined as follows [10]:

$$\text{RMSE} = \sqrt{\frac{1}{mn}\sum_{i=1}^{n}\|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2}$$

where m stands for the number of endmembers.

The following parameters impact the performance of the proposed algorithm: the regularization parameters λ and β, the augmented Lagrangian penalty parameter μ, and parameter p. We set p to 0.6 for Examples 1–3, and 0.5 for Example 4 and real data. In addition, the size K of the sliding window and the size Q of the pruned dictionary also have important effects on the performance of NCJSpLRUDP. We set the size K of the sliding window as 5×5 for all experiments. The pruned dictionary size Q is set to 10 for Example 4 and 79 for real data, as in [23]. The sliding window size K, p, and Q are discussed in Section IV-C. In all our experiments, we empirically set the penalty parameter μ as 0.1 for NCJSpLRUDP.

For SUnSAL, CLSUnSAL, SUnSAL-TV, ADSpLRU, and NCJSpLRUDP, the regularization parameters are tuned to their best respective performances in terms of varying the RMSE in $\{0, 10^{-6}, 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50\}$. In addition, for ADSpLRU, we initially set μ as 0.01 and 0.1 for all the simulated experiments and a real hyperspectral dataset, respectively. For SUnSAL-TV, we choose the optimal penalty parameter μ to vary in $\{10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1\}$. Our tests are performed on an ASUSTek laptop with an Intel(R) Core(TM) i5-5200U @ 2.20 GHz and 12.00 GB of RAM using MATLAB R2014a.

Section IV-A presents four examples used to analyze the performance of the proposed approach with a synthetic hyperspectral dataset. In Section IV-B, we qualitatively evaluate the proposed method with other methods while using a real hyperspectral dataset. Finally, we discuss the parameter selection for the proposed algorithm in Section IV-C.

A. Experiments on Simulated Datasets

Example 1 (Toy problem): In this experiment, we perform an ablation study on the effect of joint sparse and low-rank constraint to show the effectiveness of each constraint on the unmixing performance for NCJSpLRUDP algorithm. To this end, first, only the low-rankness constraint is considered by setting $\lambda = 0$ in (17). Similarly, we focus on the joint sparse constraint by setting $\beta = 0$ in (17). It should be mentioned that when one of the two parameters is fixed, the other parameter

TABLE I
RMSE AND SRE USING NCJSpLRUDP WITH DIFFERENT PARAMETER
VALUES FOR EXAMPLE 1

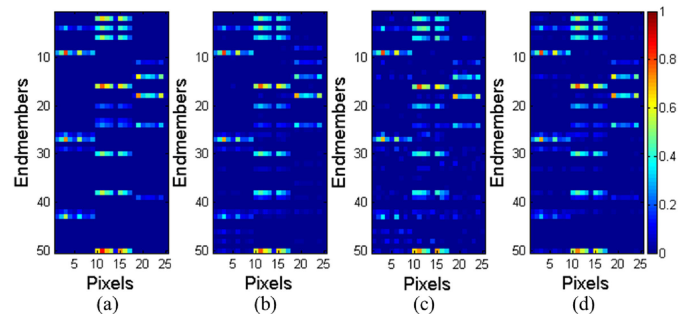| Algorithm	RMSE	SRE (dB)
NCJSpLRUDP ($\lambda = 0$)	0.0407	9.274
NCJSpLRUDP ($\beta = 0$)	0.0479	7.324
NCJSpLRUDP	0.0317	11.345

Fig. 2. Abundance maps estimated using NCJSpLRUDP with different parameter values for Example 1. (a) True abundance. (b) Abundance map for NCJSpLRUDP with $\lambda = 0$. (c) Abundance map for NCJSpLRUDP with $\beta = 0$. (d) Abundance map for NCJSpLRUDP.

is optimized when the RMSE is minimum. We randomly select $m = 50$ endmembers from $\mathbf{A}_1$ to construct our dictionary $\mathbf{A}$. Obviously, $\mathbf{A}$ is a full-rank matrix. Thus, the proposed algorithm does not use the pruning dictionary strategy, that is, $\mathbf{A}_Q = \mathbf{A}$. $K = 25$ pixels are generated via using LMM and corrupted by Gaussian noise with a signal-to-noise ratio (SNR) of 30 dB. The rank of the abundance matrix is set to 3, and the sparsity levels for three blocks are set to 15% (i.e., 15% of the entries are nonzero), 20%, and 10%. Each experiment is independently performed 30 times.

Table I lists the RMSE and SRE obtained using NCJS-pLRUDP with different parameter values. Table I indicates that NCJSpLRUDP with only the low-rankness constraint achieves better performance than NCJSpLRUDP with only the joint sparse constraint. It is again shown that both low-rank and joint sparse constraints are effective for spectral unmixing. In addition, the minimum RMSE and the highest SRE were obtained by concatenating low rank and joint sparsity. Fig. 2 shows abundance maps estimated using NCJSpLRUDP with different parameter values. We can see from Fig. 2 that NCJSpLRUDP that simultaneously considers sparsity and low rank yields more accurate sparse estimation. These results show that our method simultaneously imposing sparse and low rank is effective.

Example 2 (Different low-rankness and sparsity levels): Our purpose with this example is to show the effectiveness of the proposed algorithm for different low-rank and sparsity levels. Now, $m = 50$ endmembers are randomly extracted from $\mathbf{A}_1$ to form our dictionary $\mathbf{A}$. Similar to Example 1, $\mathbf{A}_Q = \mathbf{A}$. In this section, we design four different experiments, each of which is independently performed 30 times. For the four experiments, the abundance matrices are simultaneously considered with sparse and low ranks, and their ranks are set as either 3 or 5, with a sparsity level of either 10% or 20%, respectively. For each

TABLE II
RMSE, SRE, AND TIME USING THE CONSIDERED UNMIXING ALGORITHMS FOR EXAMPLE 2

| rank	sparsity	Criteria	SUnSAL	CLSUnSAL	SUnSAL-TV	ADSpLRU	NCJSpLRUDP
3	10%	RMSE	0.0254	0.0339	0.0246	0.0219	0.0188
		SRE(dB)	10.9315	10.0751	11.1849	13.5930	15.3763
		Time(s)	0.0197	0.2307	24.6969	0.5558	1.3291
3	20%	RMSE	0.0739	0.0706	0.0691	0.0512	0.0496
		SRE(dB)	5.4546	6.7720	6.8152	8.0825	8.8579
		Time(s)	0.0139	0.1370	24.8686	0.6174	1.4899
5	10%	RMSE	0.0339	0.0311	0.0319	0.0242	0.0211
		SRE(dB)	9.9810	10.2247	10.2139	11.9217	12.2653
		Time(s)	0.0185	0.1739	24.5052	0.4403	1.3159
5	20%	RMSE	0.0750	0.0723	0.0727	0.0657	0.0613
		SRE(dB)	5.7043	5.7609	6.2141	6.4638	6.9074
		Time(s)	0.0130	0.1441	24.6435	0.6210	1.4381
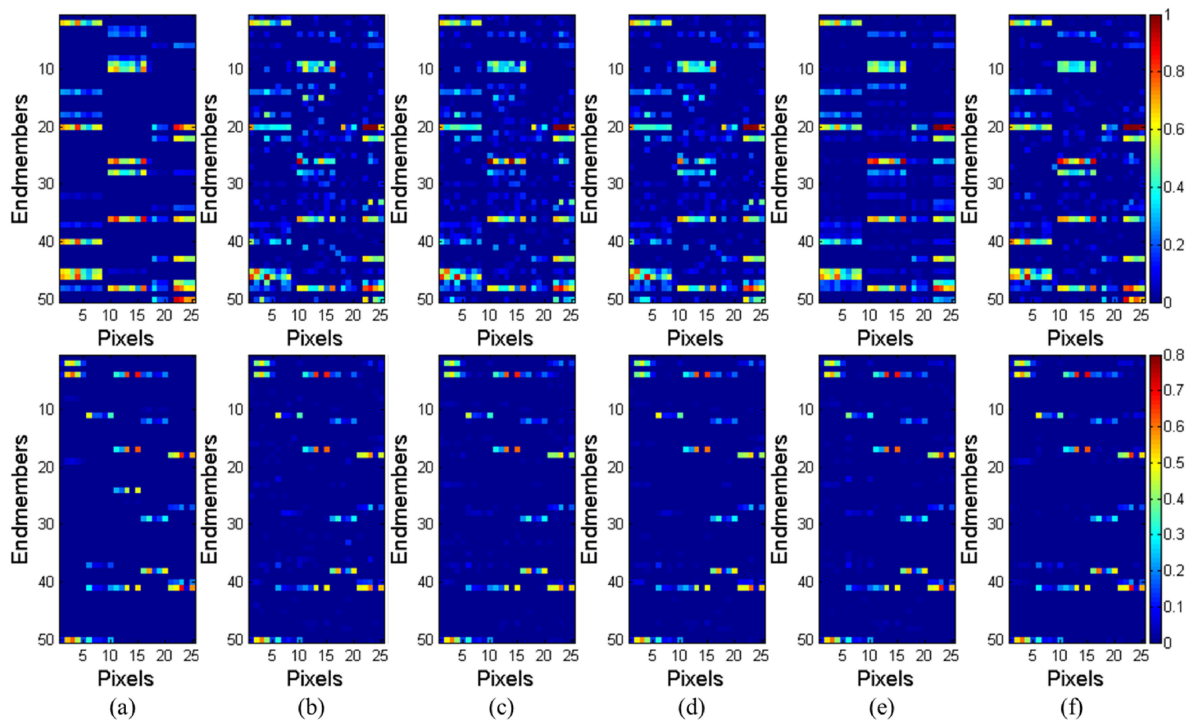
Fig. 3. Estimated abundance maps obtained with different unmixing algorithms for observation data corrupted by Gaussian noise with an SNR of 30 dB for Example 2. Top: abundance maps of rank 3 and sparsity level 20%. Bottom: abundance maps of rank 5 and sparsity level 10%. (a) True abundance. (b) Abundance maps for SUnSAL. (c) Abundance maps for CLSUnSAL. (d) Abundance maps for SUnSAL-TV. (e) Abundance maps for ADSpLRU. (f) Abundance maps for NCJSpLRUDP.

experiment, $K = 25$ pixels are generated via using LMM and corrupted by Gaussian noise with an SNR of 30 dB.

Table II lists the RMSE, SRE, and time obtained by applying SUnSAL, CLSUnSAL, SUnSAL-TV, ADSpLRU, and NCJS-pLRUDP for Example 2.

From Table II, we can see that SUnSAL-TV provides better results than SUnSAL and CLSUnSAL for the four experiments. This is because SUnSAL-TV takes into account the spatial correlation of pixels, but this feature increases the computation time. Also note that ADSpLRU and NCJSpLRUDP outperform the first three algorithms, since they are simultaneously sparse and low rank. However, NCJSpLRUDP provides better results than ADSpLRU. As per Table II, SUnSAL and CLSUnSAL are much faster than the other three algorithms, and the

proposed algorithm is faster than SUnSAL-TV but slower than ADSpLRU for all the tests. This is because the ADSpLRU algorithm executes the singular value decomposition once per iteration, while NCJSpLRUDP algorithm performs it four times. In general, NCJSpLRUDP algorithm outperforms the other methods.

Fig. 3 shows the true and estimated abundance maps obtained by the different unmixing algorithms for Example 2. We can see from Fig. 3 that SUnSAL-TV outperforms SUnSAL and CLSUnSAL because it exploits the spatial contextual information and spectral information simultaneously. Obviously, the results of NCJSpLRUDP algorithm are closer to the true abundances compared to the cases of the other four algorithms. The results in Table II are thus further corroborated.
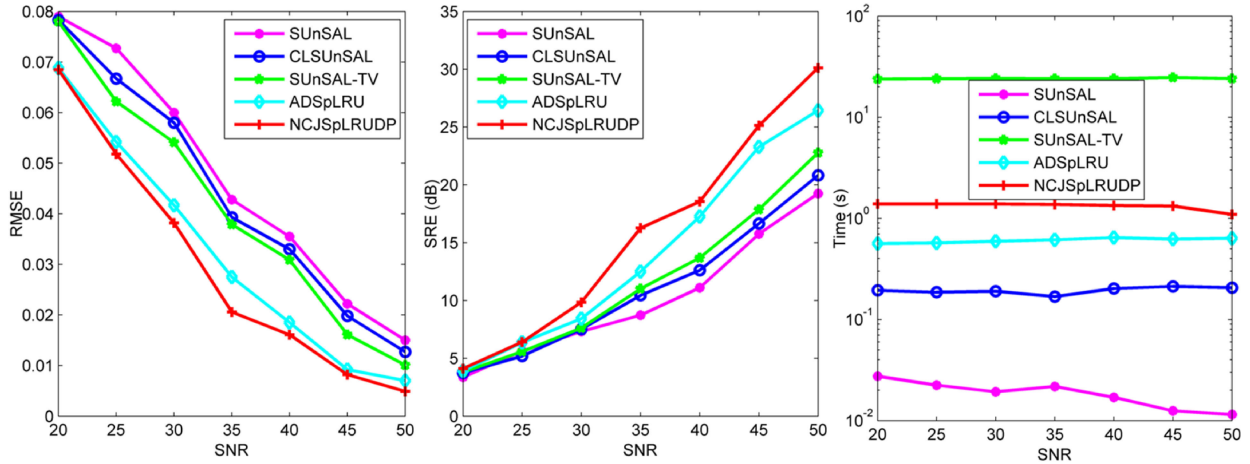
Fig. 4.　Average RMSE, SRE, and time obtained with different unmixing algorithms for different SNRs for Example 3.
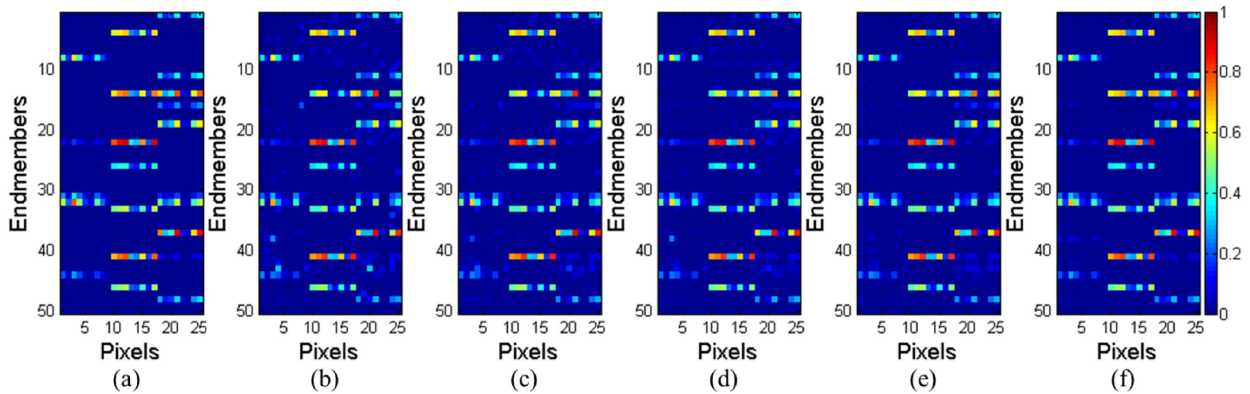

Fig. 5.　Estimated abundance maps obtained by different unmixing algorithms for the simulated data corrupted by the Gaussian noise of SNR = 35 dB for Example 3. (a) True abundance. (b) Abundance map for SUnSAL. (c) Abundance map for CLSUnSAL. (d) Abundance map for SUnSAL-TV. (e) Abundance map for ADSpLRU. (f) Abundance map for NCJSpLRUDP.

Example 3 (Different noise levels): In this experiment, we verify the performance of the proposed algorithm using simulation datasets polluted by Gaussian noise with different SNRs (ranging from 20 to 50 dB). We randomly select $m = 50$ endmembers from $\mathbf{A}_1$ to construct our dictionary $\mathbf{A}$. Similar to Example 1, $\mathbf{A}_Q = \mathbf{A}$. $K = 25$ pixels are generated via using LMM and considering pollution with Gaussian noise. The rank of each abundance matrix is set to 3, and the sparsity levels for three blocks are set to 10%, 15%, and 20%, respectively. Each experiment is independently performed 30 times.

Fig. 4 shows the average RMSE, SRE, and time obtained with different unmixing algorithms for different SNRs. The proposed algorithm provides better RMSEs and SREs than the other four algorithms for all the examined SNRs. This is because NCJS-pLRUDP algorithm considers low rank and sparsity simultaneously with dictionary pruning and thus obtains a more accurate solution. Moreover, we can observe that NCJSpLRUDP is faster than SUnSAL-TV but slower than SUnSAL, CLSUnSAL, and ADSpLRU. We can also see that the higher the SNR, the better the performance and the lower the running time of the algorithm.

Fig. 5 shows that ADSpLRU and NCJSpLRUDP provide better results than SUnSAL, CLSUnSAL, and SUnSAL-TV. Thus,

algorithms that consider sparsity and low rank simultaneously are better than the traditional unmixing algorithms. It can also be observed that the abundance matrix estimated by NCJSpLRUDP is closer to the true abundance matrix than that provided by ADSpLRU (see Fig. 5). In addition, the images for the other SNRs lead to a similar conclusion, and therefore, we omit them here.

Example 4 (Widely used simulated hyperspectral data): We illustrate the performance of the proposed algorithm using widely used simulated hyperspectral data [8], [10], [22]. The simulated data cube is generated with library $\mathbf{A}_2$ and contains 75×75 pixels with 224 banks per pixel. Each simulated pixel is generated via the LMM by five randomly selected signatures from $\mathbf{A}_2$ as the endmembers. Fig. 6(a) shows the simulated image and Fig. 6(b)–(f) shows the true fractional abundances of the five endmembers. The obtained data cubes are corrupted with the independent identically distributed (i.i.d.) Gaussian noise at three SNRs, namely, 20, 30, and 40 dB. For the fairness of the comparison experiments, RMUSIC was also used to prune the original spectrum library for the comparison methods of SUn-SAL, CLSUnSAL, SUnSAL-TV, and ADSpLRU in Example 4. The size Q of the pruned dictionary was set to 10 as in
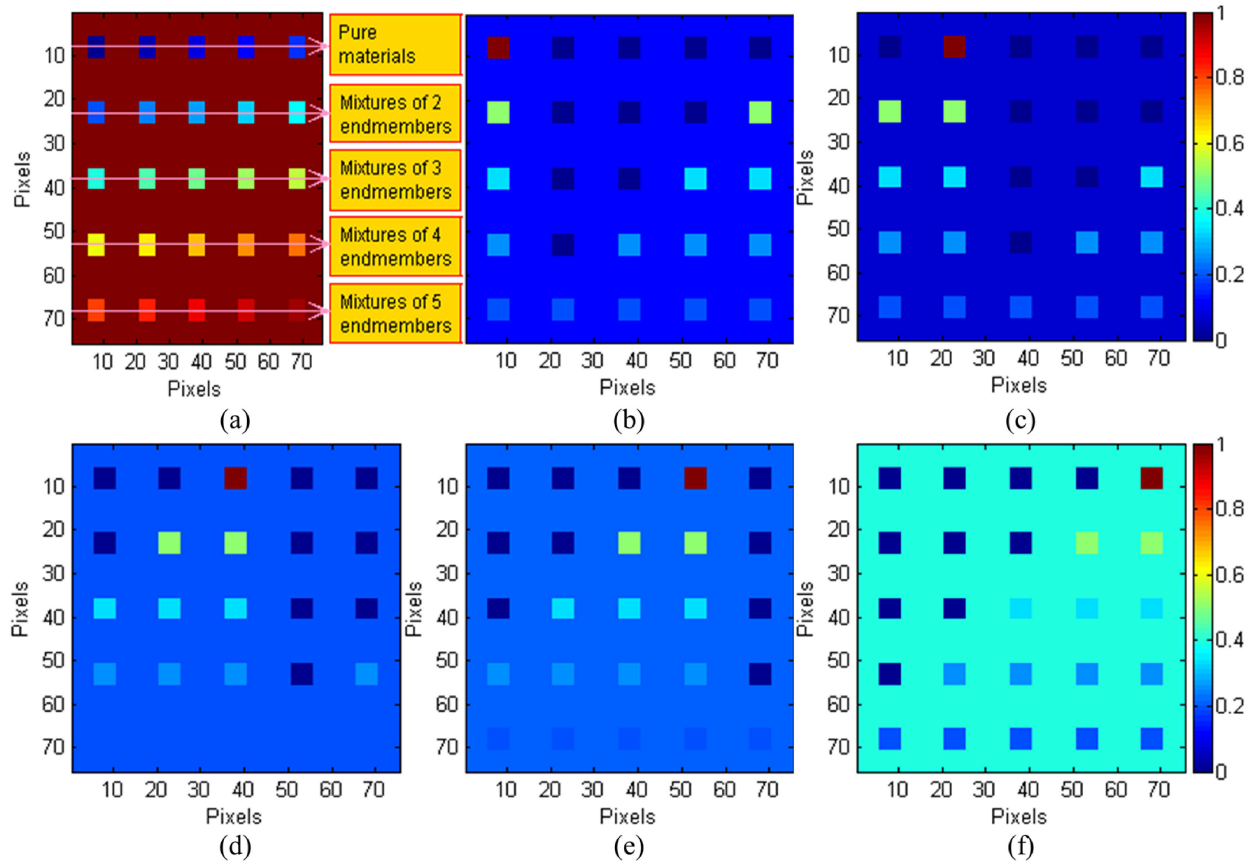
Fig. 6. True fractional abundances of endmembers for Example 4. (a) Simulated image. (b) Endmember 1. (c) Endmember 2. (d) Endmember 3. (e) Endmember 4. (f) Endmember 5.

TABLE III
RMSE, SRE, AND TIME USING THE CONSIDERED UNMIXING ALGORITHMS FOR EXAMPLE 4

SNR(dB)	Criteria	RMUSIC + SUnSAL	RMUSIC + CLSUnSAL	RMUSIC + SUnSAL-TV	RMUSIC + ADSpLRU	NCJSpLRUDP
	RMSE	0.0235	0.0116	0.0085	0.0165	0.0069
	SRE(dB)	2.6910	8.1987	11.296	6.6215	13.5811
20	Time(s)	2.6929	3.1704	127.6589	13.0887	139.2921
	Parameters	$\lambda = 0.5$	$\lambda = 5$	$\lambda = 0.001$ $\lambda_{TV} = 0.05$	$\lambda = 0.01$ $\tau = 10$	$\lambda = 5 \times 10^{-4}$ $\beta = 0.5$
	RMSE	0.0105	0.0062	0.0051	0.0082	0.0035
	SRE(dB)	9.661	14.1514	16.1499	12.6731	19.7350
30	Time(s)	2.3220	3.2792	127.1619	17.4397	117.7568
	Parameters	$\lambda = 0.05$	$\lambda = 1$	$\lambda = 0.001$ $\lambda_{TV} = 0.01$	$\lambda = 0.005$ $\tau = 0.5$	$\lambda = 5 \times 10^{-4}$ $\beta = 0.1$
	RMSE	0.0037	0.0026	0.0010	0.0036	0.0007
	SRE(dB)	19.0705	22.0265	30.3006	19.5274	33.3258
40	Time(s)	1.9557	3.2825	125.4895	18.3184	32.8915
	Parameters	$\lambda = 0.01$	$\lambda = 0.5$	$\lambda = 0.001$ $\lambda_{TV} = 0.005$	$\lambda = 0.005$ $\tau = 5$	$\lambda = 5 \times 10^{-4}$ $\beta = 0.005$

NCJSpLRUDP. We run the algorithm ten times, and average RMSE, SRE, and time are recorded. Table III lists these values obtained by implementing different unmixing algorithms for Example 4.

Table III shows that CLSUnSAL outperforms SUnSAL because it adopts the collaborative sparse strategy. Since SUnSAL-TV takes spatial information into account, it achieves better results than SUnSAL and CLSUnSAL. In addition, AD-SpLRU, which is based on low rank, provides worse results than

SUnSAL-TV because the high mutual correlation of the spectral library seriously affects the performance of the algorithm. Although NCJSpLRUDP and ADSpLRU are both low-rank and sparse unmixing models, the former achieves the best performance. This is because it adopts the dictionary pruning strategy and the sliding window. Simultaneously, we find that NCJSpLRUDP is only faster than SUnSAL-TV when the observed data have high SNR levels, because it adopts the sliding window, which increases the computational duration of the algorithm.
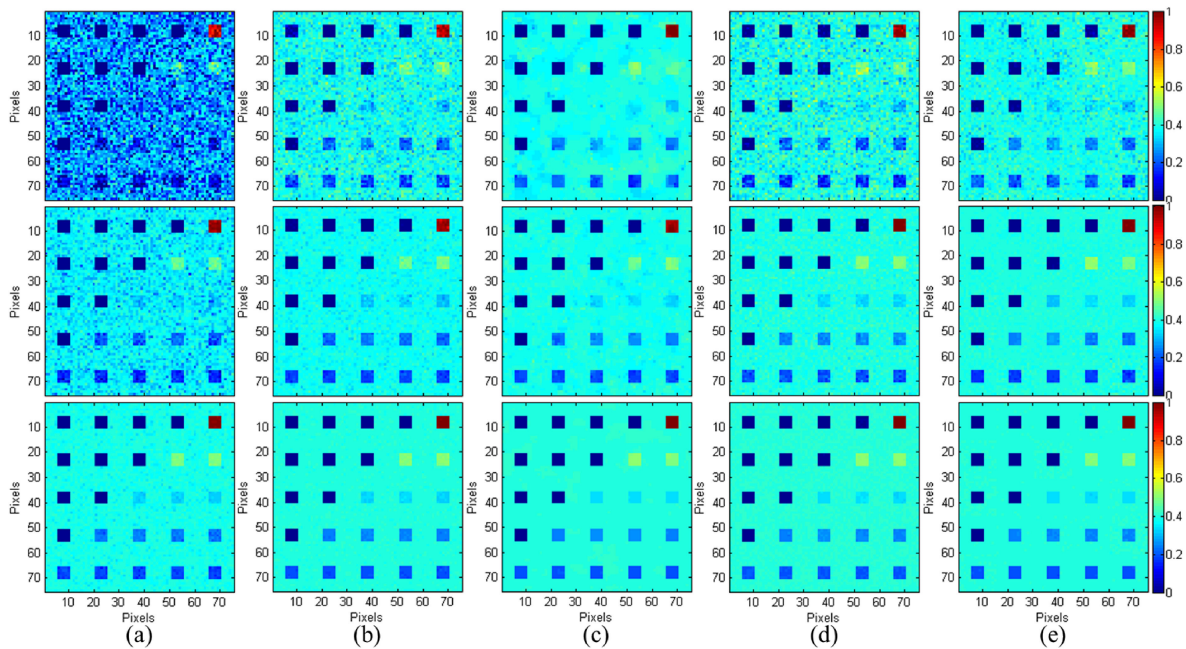
Fig. 7. Estimated abundance maps obtained by different unmixing algorithms for endmember 5 for Example 4. From top to bottom, the SNR is 20, 30, and 40 dB. (a) Abundance maps for RMUSIC + SUnSAL. (b) Abundance maps for RMUSIC + CLSUnSAL. (c) Abundance maps for RMUSIC + SUnSAL-TV. (d) Abundance maps for RMUSIC + ADSpLRU. (e) Abundance maps for NCJSpLRUDP.

Fig. 7 shows the abundance maps obtained with the different unmixing algorithms for endmember 5 for Example 4. We can see that the abundance maps obtained by SUnSAL are full of noise points, especially when obtained from images with a low SNR, and it is difficult to identify the endmember signature from the mixed spectral. Clearly, CLSUnSAL and ADSpLRU are better than SUnSAL for the three SNRs. Since the spatial information between pixels is considered, SUnSAL-TV achieves better results than SUnSAL, CLSUnSAL, and ADSpLRU. However, in some areas, the abundance maps obtained by SUnSAL-TV have an oversmooth visual effect. The NCJSpLRUDP algorithm thus provides the best results. However, we also observe that misidentification occurs in some small areas. The data presented in Table III are thus further validated.

B. Experiment on a Real Dataset

In this section, we validate the performance of the proposed algorithm on the well-known AVIRIS Cuprite dataset [8], which comprises 224 spectral bands ranging from 0.4 to 2.5 μm. A subset of the data consisting of 250×190 pixels is used for the experiments, and the number of spectral bands is reduced to 188 after removing the water absorption bands and the low SNR. Fig. 8 shows a mineral map produced in 1995 by the USGS. The USGS map serves as a good indicator for qualitative assessment of the fractional abundance maps produced by the various unmixing algorithms [6], [9], [22]. And, the scene enclosed by the red rectangle is used in the real experiment (see Fig. 8). The spectral library $\mathbf{A}_3$ used in this experiment as it includes all exposed minerals of interest.

According to [20], parameter λ for SUnSAL and CLSUnSAL is empirically set to 0.01 and 0.1, respectively. For SUnSAL-TV, λ and λ_{TV} are set to 0.001, as in [14]. In

addition, we set parameters $\lambda = \tau = 0.001$ for ADSpLRU, as in [10]. For NCJSpLRUDP, we use fine-tuned parameters $\lambda = 1 \times 10^{-4}$ and $\beta = 0.001$. Similar to Example 4, we use RMUSIC to obtain the pruned spectrum library, which is used for SUnSAL, CLSUnSAL, SUnSAL-TV, and ADSpLRU. The size Q of the pruned spectrum library is the same as NCJSpLRUDP.

Fig. 9 shows the abundance maps estimated using SUnSAL, CLSUnSAL, SUnSAL-TV, ADSpLRU, and NCJSpLRUDP algorithms for three typical minerals, namely, alunite, buddingtonite, and chalcedony. We can see that all five methods produce similar abundance maps. The abundance maps obtained by SUnSAL-TV demonstrate good spatial consistency for the typical minerals and less outliers than the maps provided by SUnSAL and CLSUnSAL because SUnSAL-TV considers the spatial information between pixels. ADSpLRU do not exhibit good spatial consistency of minerals of interest compared with SUnSAL-TV. Obviously, the NCJSpLRUDP algorithm achieves better results than the other algorithms, and the abundance maps resulting from it typically have stronger intensities for the area of interest. This is because the NCJSpLRUDP algorithm simultaneously imposes low rank and sparsity on the abundance matrix, and we also adopt dictionary pruning to eliminate the mutual correlation of the spectral library, which further improves the performance of NCJSpLRUDP.

C. Parameter Selection

In this section, we discuss the parameter selection for NCJSpLRUDP. We refer to parameter p, the sliding window size K, and the pruned dictionary size Q. We use the simulated data cube for Example 4, which was corrupted with the i.i.d. Gaussian noise with an SNR of 35 dB.
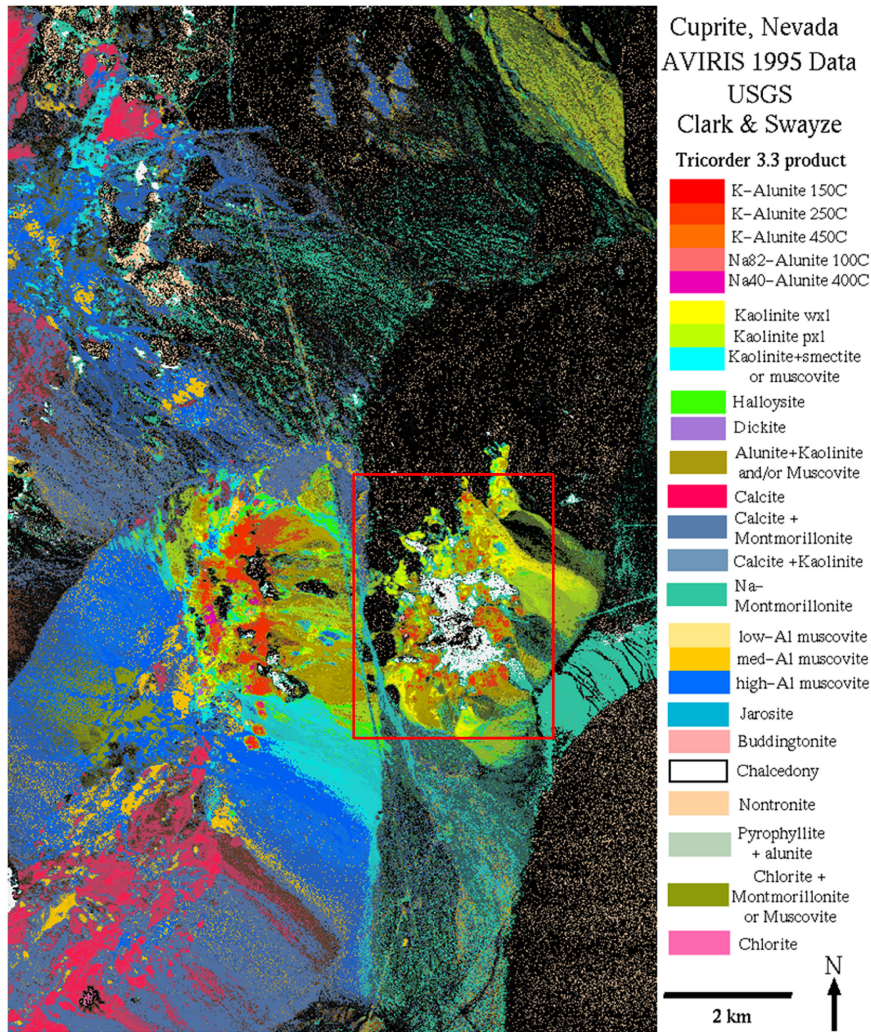
Fig. 8.   USGS map showing the locations of different minerals in the Cuprite mining district in Nevada. The map is available online at http://speclab.cr.usgs.gov/cuprite95.tgif.2.2um_map.gif. The scene enclosed by the red rectangle is used in the real experiment.

1) Selection of Parameter p: To select the optimal value of p in this experiment, we set the sliding window size to 5×5, $\mu = 0.1$, $\lambda = 5 \times 10^{-4}$, and $\beta = 0.01$ for NCJSpLRUDP. We consider ten different values of p, ranging from 0.1 to 1, and run the experiment independently ten times for each value. Fig. 10 shows the average RMSE, SRE, and time for different values of p. NCJSpLRUDP provides better performance for p values close to 0.5 and converges more quickly for small values of p. However, it is also found to be prone to local minima for small values of p. In fact, values between 0.4 and 0.6 represent a good compromise between the speed of convergence and the quality of the generated sparse solution. Thus, p was manually adjusted for the different types of data to achieve better unmixing results.

2) Sliding Window Size Selection: In general, hyperspectral images have low spatial resolution, so spatial correlation mainly occurs in small areas of the image, corresponding to a limited number of pixels [29]. In other words, as the size of the sliding window increases, the low-rank constraint of the corresponding abundance matrix usually weakens. To choose an optimal sliding window size, we compare the performance of NCJSpLRUDP using the simulated data cube for Example 4. Parameters other than the sliding window size are tuned to the respective best values. We independently run the experiment ten times for each fixed sliding window size, and the RMSE, SRE, and time for different values of the sliding widow size are reported in Table IV. We note that the NCJSpLRUDP algorithm achieves the best RMS and SRE when the sliding window size is equal to 5×5, and the computation time decreases as the sliding window size increases. This also illustrates the effectiveness of the algorithms based on low rank with a small window size.

3) Selection of Pruned Dictionary Size Q: We set the sliding window size as 5×5, $\mu = 0.1$, $\lambda = 5 \times 10^{-4}$, and $\beta = 0.01$ for NCJSpLRUDP. We run the experiment independently ten times for different values of Q (the size of the pruning dictionary). Fig. 11 shows the average RMSE, SRE, and time for different values of Q. We observe that the NCJSpLRUDP algorithm provides better performance for small values of Q, and it converges more quickly. However, the ground-truth materials may also be discarded due to the high noise. Thus, a larger value of Q is needed to avoid discarding of the true endmembers. That is,
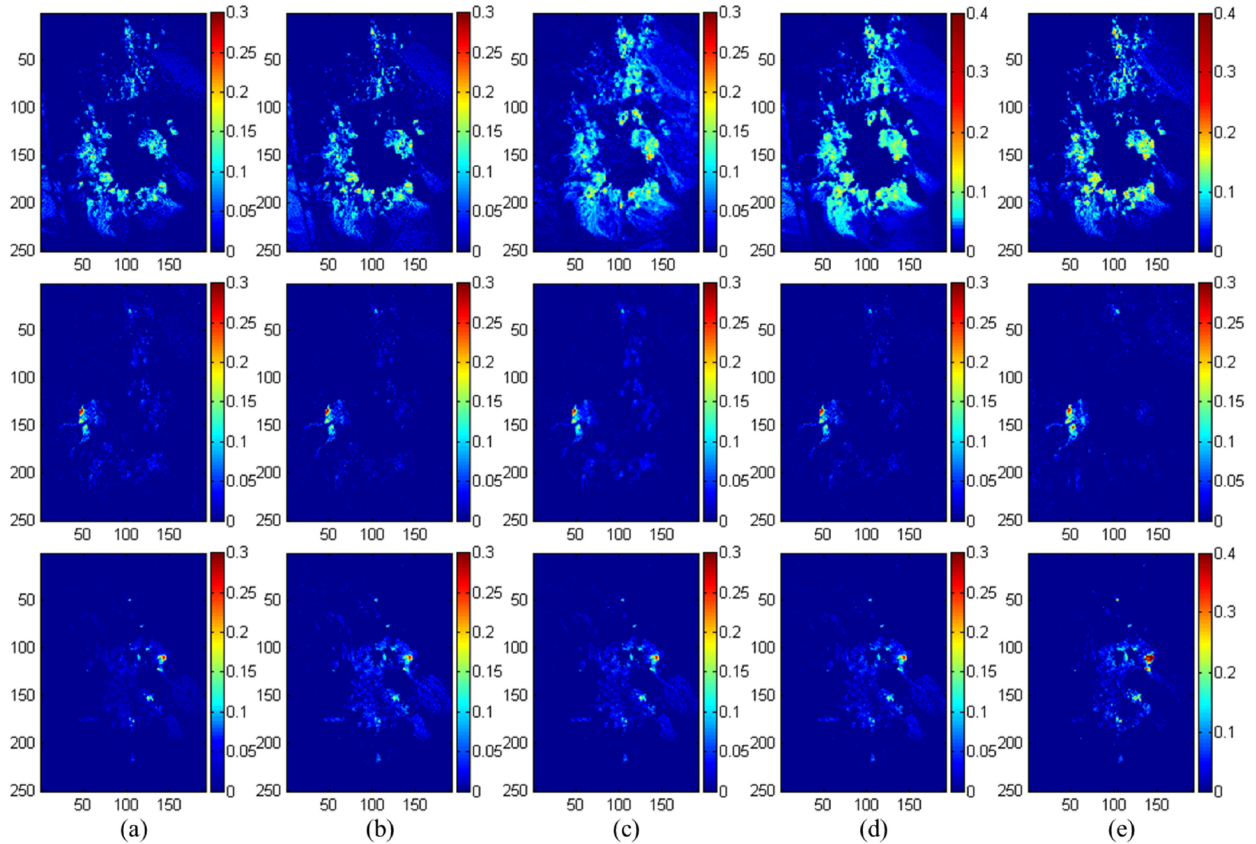
Fig. 9. Abundances maps estimated using different unmixing algorithms. Top to bottom: alunite, buddingtonite, and chalcedony. (a) Abundance maps for RMUSIC + SUnSAL. (b) Abundance maps for RMUSIC + CLSUnSAL. (c) Abundance maps for RMUSIC + SUnSAL-TV. (d) Abundance maps for RMUSIC + ADSpLRU. (e) Abundance maps for NCJSpLRUDP.

TABLE IV
PERFORMANCE OF NCJSpLRUDP FOR DIFFERENT SLIDING WINDOW SIZES FOR EXAMPLE 4

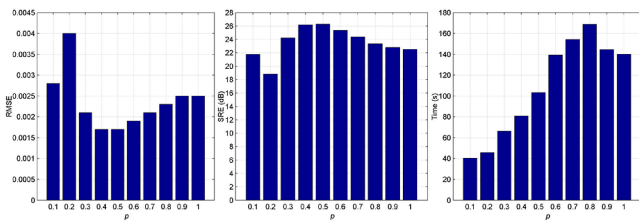| Size	3×3	5×5	7×7	9×9	11×11	13×13	15×15
RMSE	0.0059	0.0017	0.0022	0.003	0.0031	0.0034	0.0034
SRE(dB)	15.2174	26.2874	23.7941	21.192	20.8465	20.0187	19.9960
Time(s)	93.8790	103.3153	97.3008	72.4728	59.6629	56.9668	55.5751

Fig. 10. Average RMSE, SRE, and time using NCJSpLRUDP for different values of p for Example 4.
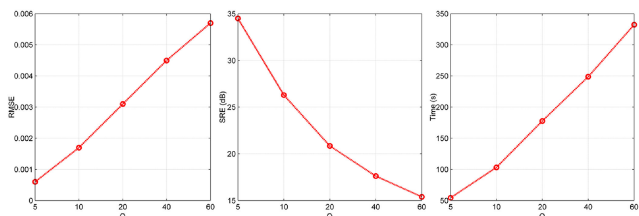

Fig. 11. Average RMSE, SRE, and time using NCJSpLRUDP for different sizes of the pruning dictionary for Example 4.

when the endmember is not in the pruned dictionary, we can increase the value of Q until all the endmembers are included in the pruned dictionary.

V. CONCLUSION

In this article, we proposed a novel model called nonconvex joint-sparsity and low-rank unmixing with dictionary pruning. The proposed method imposes sparsity on the abundance matrix via the $\ell_{2,p}$ mixed norm, which has better sparsity than $\ell_{2,1}$. At the same time, the global structure of the data is obtained using an LRR. In addition, using the dictionary pruning strategy allows us to obtain a subset of the original spectral library for our unmixing model, which eliminates the high mutual coherence of the original spectral libraries. Thus, this strategy not only ensures the effectiveness of our model, but also significantly improves the performance of the proposed method. This model can be solved with the well-known ADMM, and the simulated and real data experiments demonstrate the efficacy of the

proposed algorithm. Although highly promising, the proposed method can be further improved. For instance, the value of p and the size of the sliding window in the proposed algorithm need to be manually adjusted depending on the different types of data, which limits further applications of our algorithm. Therefore, in our future work, we plan to explore a method for the automatic selection of these parameters, thereby improving our method. Moreover, as tensor-based models can naturally maintain the spectral and spatial information of hyperspectral image data, we also plan to evaluate tensor-based unmixing methods.

REFERENCES

[1] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[2] L. Sun, Z. Wu, L. Xiao, J. Liu, Z. Wei, and F. Dang, "A novel $L_{1/2}$ sparse regression method for hyperspectral unmixing," *Int. J. Remote Sens.*, vol. 34, no. 20, pp. 6983–7001, Oct. 2013.

[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] A. Erturk, M. D. Iordache, and A. Plaza, "Sparse unmixing with dictionary pruning for hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 321–330, Jan. 2017.

[5] D. Wang, Z. Shi, and X. Cui, "Robust sparse unmixing for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1348–1359, Mar. 2018.

[6] 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.

[7] J. Sigurdsson, M. O. Ulfarsson, J. R. Sveinsson, and J. M. Bioucas-Dias, "Sparse distributed multitemporal hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6069–6084, Nov. 2017.

[8] 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.

[9] J. Liu, C. Zhang, J. Zhang, H. Li, and Y. Gao, "Manifold regularization for sparse unmixing of hyperspectral images," *SpringerPlus*, vol. 5, no. 1, Nov. 2016, Art. no. 2007.

[10] J. Huang, T. Z. Huang, L. J. Deng, and X. L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2419–2438, Apr. 2019.

[11] J. M. Bioucas-Dias, "A variable splitting augmented Lagarangian approach to linear spectral unmixing," *in Proc. 1st Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Grenoble, France, Aug. 2009, pp. 1–4.

[12] F. Yin, "Study on the theory and method of linear hyperspectral unmixing and its application," Ph.D. dissertation, Dept. Ear. Explor. Info. Technol., Chengdu Univ. Technol., Chengdu, China, 2016.

[13] Y. Zhong, R. Feng, and L. Zhang, "Non-local sparse unmixing for hyperspectral remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1889–1909, Jun. 2014.

[14] C. Zheng, H. Li, Q. Wang, and C. L. P. Chen, "Reweighted sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 479–488, Jan. 2015.

[15] X. Li, J. Huang, L. Deng, and T. Z. Huang, "Bilateral filter based total variation regularization for sparse hyperspectral image unmixing," *Inf. Sci.*, vol. 504, pp. 334–353, Jul. 2019.

[16] 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.

[17] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using ℓ_0 norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5495–5508, Sep. 2018.

[18] A. Halimi, J. M. Bioucas-Dias, N. Dobigeon, G. S. Buller, and S. McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 146–159, Jun. 2017.

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

[20] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[21] J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," *in Proc. 2nd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.

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

[23] 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.

[24] J. Miao, X. Zhou, T. Huang, T. Zhang, and Z. Zhou, "A novel inpainting algorithm for recovering Landsat-7 ETM+ SLC-OFF images based on the low-rank approximate regularization method of dictionary learning with nonlocal and nonconvex models," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6741–6754, 2019.

[25] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[26] J. Chen and J. Yang, "Robust subspace segmentation via low-rank representation," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1432–1445, Aug. 2014.

[27] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 2862–2869.

[28] Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, Jul. 2014.

[29] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4775–4789, Aug. 2016.

[30] Q. Jin *et al.*, "Gaussian mixture model for hyperspectral unmixing with low-rank representation," *in Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, Jul. 2019, pp. 294–297.

[31] H. Li, R. Feng, L. Wang, Y. Zhong, and L. Zhang, "Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.2994260.

[32] Y. Xu, "Sparse abundance estimation with low-rank reconstruction for hyperspectral unmixing," *Int. J. Remote Sens.*, vol. 41, no. 17, pp. 6805–6830, Jun. 2020.

[33] Y. Zheng, F. Wu, H. J. Shim, and L. Sun, "Sparse unmixing for hyperspectral image with nonlocal low-rank prior," *Remote Sens.*, vol. 11, no. 24, Dec. 2019, Art. no. 2897.

[34] L. Sun, F. Wu, T. Zhan, W. Liu, J. Wang, and B. Jeon, "Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1174–1188, 2020.

[35] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[36] J. Miao, X. Zhou, and T. Huang, "Local segmentation of images using an improved fuzzy C-means clustering algorithm based on self-adaptive dictionary learning," *Appl. Soft Comput.*, vol. 91, Jun. 2020, Art. no. 106200.

[37] Y. Xie, S. Gu, W. Zuo, W. Zhang, and L. Zhang, "Weighted Schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[38] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[39] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," *in Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 217–224.

[40] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Nov. 2007.

[41] F. Cao, M. Cai, Y. Tan, and J. Zhao, "Image super-resolution via adaptive $\ell_p(0 < p < 1)$ regularization and sparse representation," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 27, no. 7, pp. 1550–1561, Jul. 2016.

[42] L. Wang, S. Chen, and Y. Wang, "A unified algorithm for mixed $\ell_{2,p}$-minimizations and its application in feature selection," *Comput. Optim. Appl.*, vol. 58, pp. 409–421, Mar. 2014.

[43] X. Chen, F. Xu, and Y. Ye, "Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2832–2852, Feb. 2010.

[44] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

Si Guo was born in Chengdu, Sichuan, China, in 1987. She received the B.S. degree in geophysics in 2009 and the M.S. and Ph.D. degrees in sedimentology, in 2011 and 2015, respectively, all from the Chengdu University of Technology, Chengdu, China.

Since 2009, she has been an Engineer with the Reservoir Prediction Unit, CNPC Chuanqing Drilling Engineering Company, Chengdu.

Hongwei Han received the B.S. degree in mathematics education from Yan'an University, Yan'an, China, in 2002, and the M.S. degree in electronics and communication engineering, in 2009 from the Chengdu University of Technology, Chengdu, China, where he is currently working toward the Ph.D. degree in geomathematics.

His research interests include hyperspectral remote sensing image processing and partial differential equation, and its applications.

Ling Chen received the B.S. degree in mathematics education from Southwest Normal University, Chongqing, China, in 1992, and the M.S. degree in applied mathematics, in 2003, and the Ph.D. degree in geodetection and information technology, in 2011, both from the Chengdu University of Technology, Chengdu, China.

She is currently a Professor with the Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology. Her research interests include remote sensing image processing, earth exploration, and information technology.

Guxi Wang was born in Chengdu, Sichuan, China, in 1985. He received the B.S. and M.S. degrees in resources and environment from Sichuan Agricultural University, Ya'an, China, in 2009 and 2012, respectively, and the Ph.D. degree in earth exploration and information techniques from the Chengdu University of Technology, Chengdu, China, in 2017.

Since 2015, he has been a Research Associate with the National Institute of Measurement and Testing Technology, Chengdu. His research interests include image processing and environmental monitoring and analysis.

Mingyue Zhang received the B.S. degree in mathematics from Yibin University, Yibin, China, in 2016, and the M.S. degree in mathematics, in 2019 from the Chengdu University of Technology, Chengdu, China, where she is currently working toward the Ph.D. degree in geomathematics.

Her research interests include hyperspectral remote sensing geological applications.

Maozhi Wang received the B.S. degree in mathematics from Southwest Normal University, Chongqing, China, in 1996, the M.S. degree in computer software and theory from the University of Electronics Science and Technology of China, Chengdu, China, in 2003, and the Ph.D. degree in geodetection and information technology from the Chengdu University of Technology, Chengdu, in 2014.

He is currently a Researcher with the Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology. His research interests include hyperspectral remote sensing image processing and its applications.

Ke Guo received the B.S. degree in mathematics from Chengdu Geological College, Chengdu, China, in 1985, the M.S. degree in geomathematics from Chengdu Geological College, Chengdu, China, in 1990, and the Ph.D. degree in mineral resource prospecting and exploration from the Chengdu University of Technology, Chengdu, China, in 2005.

He is a Professor and Doctoral Supervisor with the Chengdu University of Technology, Chengdu, China. He is an Academic and Technical Leader of Sichuan province and enjoy special subsidies from the State Council Government. He is currently the Head of the teaching team of mathematics geology in Sichuan province, and the Head of the scientific research and innovation team of high-level resources and environment in Sichuan province. He has long been involved in the teaching and scientific research of quantitative evaluation and prediction of resources and environment.

Jiaqing Miao received the B.S. degree in mathematics from Northeast Petroleum University, Daqing, China, in 2005, and the M.S. degree from the School of Mathematics, Northeastern University, Shenyang, China, in 2008. He is currently working toward the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

He has been a Visiting Scholar with the Department of Geophysical Engineering, Montana Technological University, Butte, MT, USA, since November 2016. His current research interests include the digital image processing and applications of medical and remote sensing, including image segmentation, image inpainting, image clustering compressed sensing, and remotely sensed image reconstruction.