Abundance Estimation for Bilinear Mixture Models via Joint Sparse and Low-Rank Representation

Qing Qu, *Student Member, IEEE*, Nasser M. Nasrabadi, *Fellow, IEEE*, and Trac D. Tran, *Senior Member, IEEE*

***Abstract*—Sparsity-based unmixing algorithms, exploiting the sparseness property of the abundances, have recently been proposed with promising performances. However, these algorithms are developed for the linear mixture model (LMM), which cannot effectively handle the nonlinear effects. In this paper, we extend the current sparse regression methods for the LMM to bilinear mixture models (BMMs), where the BMMs introduce additional bilinear terms in the LMM in order to model second-order photon scattering effects. To solve the abundance estimation problem for the BMMs, we propose to perform a sparsity-based abundance estimation by using two dictionaries: a linear dictionary containing all the pure endmembers and a bilinear dictionary consisting of all the possible second-order endmember interaction components. Then, the abundance values can be estimated from the sparse codes associated with the linear dictionary. Moreover, to exploit the spatial data structure where the adjacent pixels are usually homogeneous and are often mixtures of the same materials, we first employ the joint-sparsity (row-sparsity) model to enforce structured sparsity on the abundance coefficients. However, the joint-sparsity model is often a strict assumption, which might cause some *aliasing artifacts* for the pixels that lie on the boundaries of different materials. To deal with this problem, the low-rank-representation model, which seeks the lowest rank representation of the data, is further introduced to better capture the spatial data structure. Our simulation results demonstrate that the proposed algorithms provide much enhanced performance compared with state-of-the-art algorithms.**

***Index Terms*—Abundance estimation, bilinear model, hyperspectral imagery, low-rank representation (LRR), spectral unmixing (SU).**

I. Introduction

HYPERSPECTRAL imaging (HSI) has received considerable attention in the last few decades. With the wealth of spectral information available, HSI has been successfully applied to various domains such as agriculture, mineralogy, and environment monitoring. However, because of the low spatial resolution of current HSI sensors, several distinct pure materials can jointly occupy a single pixel, causing critical problems in the accurate interpretation of the image contents. Thus, the spectral unmixing (SU) problem, which consists of identifying the pure materials (endmembers) and estimating their associated fractions (abundances), has become a major issue in HSI and has recently been extensively investigated [1], [2].

Solving the SU problem usually consists of two major steps: 1) the pure endmember extraction and 2) the abundance estimation. In the first step, all the possible pure endmembers should be extracted from the hyperspectral image [3]. Many endmember detection algorithms extract the pure endmembers under the pixel purity assumption that at least one pure pixel exists for each endmember in the scene. These algorithms include the pixel purity index [4], successive projection algorithm [5], and vertex component analysis (VCA) [6]. Some researchers have focused on the convex geometry of the hyperspectral data that all the pixels lie in a convex hull in a high-dimensional subspace. These studies include the simplex volume minimization [7], [8], [9] and simplex volume maximization techniques such as N-Finder [10], [11]. More recently, researchers try to find the endmember signatures even if there are no pure pixels existing by methods such as the nonnegative matrix factorization [12], [13]. In our proposed abundance estimation method, we assume that the pure endmembers are known or they can be identified from the target image by one of these endmember detection techniques. The second step consists of estimating the corresponding abundances under some physical constraints. In this step, the inherent source separation problem should be efficiently solved under certain assumptions on the mixture model [1], [2].

In the literature, the linear mixture model (LMM) has been widely applied for abundance estimation due to its simplicity and analytically tractable solutions. It assumes that the spectral response of a pixel is a linear combination of all the pure endmembers present in that pixel. To be physically meaningful, it usually follows two constraints: the so-called abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC) [16]. Many different algorithms have been proposed in the literature to estimate the abundances for the LMM. These algorithms are based on the least squares principle [16], geometric formulation [17], and statistical inference [18], [19].

Nonetheless, as pointed out in [1] and [2], due to the wide existence of nonlinear effects caused by the interactions of photons with multiple components in the image, the LMM may not be an appropriate model in many practical situations. Instead, nonlinear mixture models provide an interesting alternative for overcoming these inherent limitations. For instance, in the case of intimate mixtures where there are microscopic

Manuscript received November 27, 2012; revised May 9, 2013, July 9, 2013, and September 5, 2013; accepted September 7, 2013. Date of publication October 3, 2013; date of current version March 3, 2014. This work was supported in part by the National Science Foundation under Grant CCF-1117545, by the Army Research Office under Grant 60219-MA, and by Office of Naval Research under Grant N00014-12-1-0765.

Q. Qu and T. D. Tran are with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: qqu2@jhu.edu; trac@jhu.edu).

N. M. Nasrabadi is with the U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: nnasraba@arl.army.mil).

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

Digital Object Identifier 10.1109/TGRS.2013.2281981

photon interactions within the mixed materials, Hapke [20] has demonstrated that the reflectance is a nonlinear function of the apparent albedo. For multilayered mixtures where the photons are scattered from one material into another before reaching the sensor, various bilinear mixture models (BMMs) have been introduced to capture such nonlinearities [21], [22]. For example, the generalized bilinear model (GBM), recently studied in [23], generalizes the LMM by introducing bilinear terms that take into account the multipath effects. Most recently, the postnonlinear mixture model (PNMM) for source separation problems [24] has also been introduced for modeling the nonlinearities in hyperspectral images [25]. Based on different nonlinear model assumptions, various abundance estimation algorithms have been proposed such as least squares [25], kernel-based least squares [26], [27], and Bayesian methods [23].

However, none of these linear and nonlinear abundance estimation algorithms has taken into account the sparsity constraint. Sparseness is an important property of the hyperspectral imagery [13]–[15]. In most cases, every endmember does not contribute to all the pixels in the scene. In other words, the abundance of each endmember is localized, and each pixel is a mixture of a few pure endmembers from the scene. To exploit this property, in [28], Guo *et al.* have proposed a supervised l_1-regression algorithm for hyperspectral unmixing with a known small-size dictionary. Similar to this method, most recently, Iordache *et al.* [29] proposed a semisupervised sparse unmixing algorithm which uses a selected overcomplete U.S. Geological Survey (USGS) spectral library.[1] The aforementioned papers both take the sparsity property of abundances into account and solve an l_1-minimization problem using a dictionary made up of pure endmembers. Because of the l_1-norm regularization, they produce sparser and more stable solutions than state-of-the-art algorithms. However, these methods have their own problems and limitations.

1) Because the endmembers are usually correlated, using a huge overcomplete dictionary (around 500 pure endmembers) with a high mutual coherence might result in unsatisfactory unmixing performances which was shown in [29], let alone the computational burden due to the large dictionary size.
2) These sparsity-based algorithms are only dedicated to the LMM, which cannot effectively deal with the nonlinearities caused by the multilayer light scattering effects.

With these considerations, we assume that the endmember dictionary is known and of small size, which can be obtained by an endmember extraction technique. Based on this assumption, we propose a novel sparse regression-based approach to estimate the abundances for the BMM. The BMM accounts for the multiple light scattering effects by introducing an additional bilinear term to the LMM. Therefore, we extend the linear sparse regression method in [28] and [29] to the BMM by using an additional bilinear dictionary, which takes into account all the possible second-order bilinear interaction components of the pure endmembers. We concatenate two dictionaries—the original linear dictionary, which contains all the pure endmembers, as well as the bilinear dictionary representing all the bilinear components to form a so-called composite dictionary. Since the components within the bilinear term are also linearly combined, the overall problem still remains as a simple linear sparse regression task based on this composite dictionary, which transforms the bilinear problem into a linear problem. The new sparse regression problem can be efficiently solved via the l_1-minimization techniques. Among all the optimization methods, we adopt the nonnegative constraint alternating direction method of multipliers (ADMM) [30], which has been demonstrated for its simplicity, efficiency, and robustness [31]. Once the sparse representation for each mixed pixel has been obtained, the endmembers are identified, and their corresponding abundances are immediately estimated from the sparse codes associated with the linear dictionary.

Furthermore, recent research [32]–[36] has demonstrated that the spatial information, which characterizes the relationship between each pixel vector and its neighbors, can also be incorporated into the unmixing algorithms to enhance the performance of the abundance estimation. Specifically, with the assumption that the abundance is piecewise smooth, Iordache *et al.* [32] minimize the difference of abundances between adjacent pixels by the total variation (TV) constraint. The TV regularization requires a rather strict assumption that the abundance of neighboring pixels is piecewise smooth, which means that both the mixing material and its associated abundance should be similar for adjacent pixels. The Laplacian graph regularization [35], [36] minimizes the weighted difference of two neighboring abundance vectors when the signatures of the pixels are similar, by the assumption that, if two data points are similar to each other, then their abundance representations with respect to (w.r.t.) the dictionary should also be similar. Iordache *et al.* [33] have introduced the joint-sparsity constraint [37] to hyperspectral unmixing enforcing neighboring pixels to have the same endmembers, due to the observation that adjacent pixels usually consist of very similar endmembers in the homogeneous region. Compared with the TV regularization, the joint-sparsity constraint is less strict. It assumes that neighboring pixels are made of similar materials but do not necessarily have similar abundance for each material, which is a more general assumption in the HSI.

Therefore, we use the joint-sparsity constraint to exploit the spatial information for the bilinear abundance estimation problem. To use the spatial structured sparsity in the data, instead of recovering the abundance representation for each single measurement vector (SMV) individually, we consider the multiple measurement vector (MMV) problem [38] for a joint-sparse regression, which enforces the sparse representations for all the neighboring pixels to share the same support set. In our proposed method, we enforce the joint-sparsity structure on each target pixel and all of its eight neighbors in a 3×3 sliding window. We slide the window through the whole image one pixel at a time to estimate the abundance for every pixel in the center of the window. To solve the joint-sparse regression problem, we introduce a nonnegative constrained MMV-ADMM algorithm which minimizes the $l_{1,2}$-norm of the abundance matrix that contains all the abundance vectors of the pixels within the sliding window. Because the joint-sparsity

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

model exploits both the spectral and the contextual information, it further improves the abundance estimation performance.

Further investigation shows that the joint-sparsity model might cause some undesirable misrepresentation artifacts (called *aliasing artifacts* in this paper) for the pixels on the boundaries between different regions when the materials are distinct (see Fig. 2). In HSI, the high spatial similarity usually implies the low column rank property of the data. This property further indicates that the abundance representation matrix has a low-rank characteristic rather than the strict row sparsity. Recently, Liu *et al.* [39], [40] have proposed a novel low-rank representation (LRR) technique for subspace clustering problems, which jointly obtains a representation for all the data under a global lowest rank constraint. To handle the outliers in the data, they further introduce a sparse error matrix to the model. In this paper, we modified the original LRR model in [40] to solve the BMM-based abundance estimation problem. First, the bilinear terms in BMMs are "analogous" to the outliers in LRR, so we model the bilinear terms by a bilinear matrix, factorizing it into a bilinear dictionary and its associated bilinear representation. Second, because abundances are required to satisfy the ASC and ANC, we enforce both of these constraints on the proposed model. As the modified LRR model better captures the spatial data structure than the joint-sparsity model, it provides better abundance estimation results compared with state-of-the-art techniques.

In summary, our main contributions in this paper are twofold.

1) We introduce a simple but very effective sparsity-based abundance estimation method for BMMs. Because all the components in BMMs are linearly combined, we propose to transform the bilinear problem into a linear one by introducing a composite dictionary consisting of a linear and a bilinear dictionary. The proposed method shows much better estimation results than state-of-the-art algorithms for BMMs.
2) We propose a generalized LRR model for the abundance estimation problem which further exploits the spatial information in HSI. Because the high spatial correlation of the data implies the low-rank property of the abundance matrix, the proposed LRR model can better capture the spatial data structure by seeking the lowest rank representation, which shows much enhanced performance in turn.

The rest of this paper is structured as follows. The LMM and several BMMs are summarized in Section II. The new sparse regression algorithm for BMMs is proposed in Section III. In Section IV, the joint-sparsity model and its optimization method are introduced. A modified LRR model is then proposed in Section V. In Section VI, the effectiveness of the proposed method is demonstrated by various experiments on both synthetic and real data. Finally, we conclude this paper with a few remarks in Section VII.

II. Mixture Models

Suppose that $\mathbf{y} = [y_1, y_2, \ldots, y_L]^T$ is an observed mixed pixel of R pure endmembers with L spectral bands. We assume that the dictionary $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_R] \in \mathbb{R}^{L\times R}$ in which each column $\mathbf{a}_r = [a_{1,r}, a_{2,r}, \ldots, a_{L,r}]^T (1 \le r \le R)$ is a pure endmember vector. Let $\mathbf{x} = [x_1, x_2, \ldots, x_R]^T$ be an abundance vector associated with the observed pixel $\mathbf{y}$.

A. LMM

The underlying physical assumption of the LMM is that each incident photon interacts with only one earth surface component and that the reflected spectra are not scattered and therefore do not mix before entering the HSI sensor. In this case, the model can be described as

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

where $\mathbf{n} = [n_1, n_2, \ldots, n_L]^T$ is an additive white noise sequence, which is usually assumed as an independent and identically distributed zero-mean Gaussian sequence with variance σ^2, denoted as $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. To be physically meaningful, the abundance vector $\mathbf{x}$ has to be nonnegative, and the summation of all its fractions is supposed to be one. Thus, it should satisfy two constraints, i.e., ANC and ASC, respectively

$$\begin{aligned} &\text{ANC}: x_r \ge 0, \forall r \in 1, 2, \ldots, R \\ &\text{ASC}: \sum_{r=1}^{R} x_r = 1. \end{aligned} \tag{2}$$

The abundance coefficient vectors in all the following nonlinear models are supposed to satisfy the two constraints in (2) as well.

B. Generalized BMM

To account for the presence of multiple photon bounces, the BMM introduces an additional bilinear term to the LMM. Specifically, the GBM [23] assumes that the observed pixel can be expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{ij} x_i x_j \mathbf{a}_i \odot \mathbf{a}_j + \mathbf{n} \tag{3}$$

where $\gamma_{ij} (1 \le i \le j \le R)$ is a nonnegative coefficient that controls the interaction between the ith and jth endmembers in the mixed pixel $\mathbf{y}$. Because a backscattered path is usually longer than the corresponding direct reflection, the parameter γ_{ij} is constrained by

$$0 \le \gamma_{ij} \le 1, i, j = 1, \ldots, R, i \ne j$$

where the symbol $\odot$ denotes the Hadamard product operation

$$\mathbf{a}_i \odot \mathbf{a}_j = \begin{pmatrix} a_{1,i} \\ \vdots \\ a_{L,i} \end{pmatrix} \odot \begin{pmatrix} a_{1,j} \\ \vdots \\ a_{L,j} \end{pmatrix} = \begin{pmatrix} a_{1,i}a_{1,j} \\ \vdots \\ a_{L,i}a_{L,j} \end{pmatrix}. \tag{4}$$

From (2), we can see that $0 \le x_k \le 1 (\forall 1 \le k \le R)$. Therefore, $\gamma_{ij} x_i x_j$ is usually much smaller than unity. Thus, in GBM, the total contribution of the bilinear components is small, and

the linear part dominates the mixture. Furthermore, note that GBM reduces to the Fan's bilinear model (FM) studied in [42] when $\gamma_{ij} = 1(1 \leq i, j \leq R, i \neq j)$, and it degenerates to LMM when $\gamma_{ij} = 0(1 \leq i, j \leq R, i \neq j)$.

C. PNMM

The PNMM was recently proposed in [25] for modeling the hyperspectral nonlinear mixtures. It can be described as

$$\mathbf{y} = g(\mathbf{Ax}) + \mathbf{n} \tag{5}$$

where $g(\cdot)$ should be an appropriate nonlinear function mapping from $[0,1]^L$ into $[0,1]^L$. Various functions such as polynomial and sigmoidal functions have shown the potential for modeling the nonlinear mixtures. In [25], the authors consider the second-order polynomial nonlinearity as

$$g : [0,1]^L \to [0,1]^L$$
$$\mathbf{s} \to \left[s_1 + bs_1^2, \ldots, s_L + bs_L^2\right]^T$$

where $\mathbf{s} = [s_1, \cdots, s_L]^T$ and b is a parameter balancing the nonlinear effects. As stated in [25], the particular choice of second-order polynomial has the advantage of defining the nonlinearity by a unique parameter b, whose value allows the importance of the nonlinear terms to be characterized. Another motivation is the Weierstrass approximation theorem, which states that every continuous function defined on an interval can be uniformly approximated by a polynomial with any desired precision [41]. Thus, they propose a polynomial PNMM (PPNMM) to model the nonlinearities as

$$\mathbf{y} = \mathbf{Ax} + b(\mathbf{Ax}) \odot (\mathbf{Ax}) + \mathbf{n}. \tag{6}$$

The PPNMM differs from the previously introduced GBM in that it considers not only the interaction products between endmembers but also the self-products of each endmember. Hence, it models the scattering among different endmembers as well as within each material (see Fig. 1). However, although the single parameter b makes PPNMM much easier to estimate, it might not be as flexible as GBM in characterizing the strength of the contributions for different bilinear components.

D. MGBM

From (3) and (6), we can see that both the GBM and PPNMM have their inherent limitations.

1) The PPNMM differs from the previously introduced GBM in that it considers not only the interaction products between endmembers but also the self-products of each endmember. Hence, it can model the scattering effects within each material, where GBM cannot (see Fig. 1). However, although the single parameter b makes PPNMM much easier to estimate, it might not be as flexible as GBM in characterizing the strength of contributions for different bilinear components.
2) Both GBM and PPNMM make the assumption that a nonlinearly mixed pixel only contains bilinear terms involving the endmembers which have nonzero linear abundance. However, in some scenarios, such an assumption might not hold. For instance, the reflections of materials (i.e., buildings) from a neighboring pixel might interact with the materials in the target pixel and therefore contribute to its bilinear terms. However, since the materials are not located in the target pixel, it will have a zero linear abundance. However, neither GBM nor PPNMM can handle such cases.

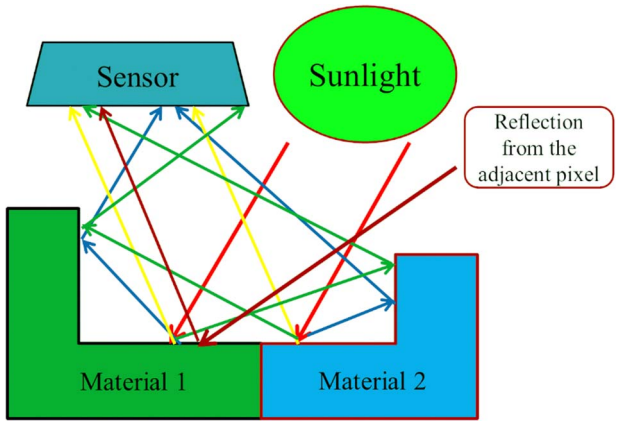


Fig. 1. Second-order multipath scattering effects in nonlinear mixing. The red and yellow lines denote the incident and the directly reflected light, respectively. The blue and green lines indicates the backscattered lights within and among the materials in order. The black line is the light from the reflection from the adjacent pixel.

To address these limitations, we consider a modified GBM (MGBM) as follows:

$$\mathbf{y} = \mathbf{Ax} + \sum_{i=1}^{R}\sum_{j=i}^{R} \zeta_{ij}\mathbf{a}_i \odot \mathbf{a}_j + \mathbf{n} \tag{7}$$

where $0 \leq \zeta_{ij} \ll 1(1 \leq i, j \leq R)$. The proposed model in (7) is a more general model compared with GBM and PPNMM: 1) It considers the polynomial terms $\mathbf{a}_i \odot \mathbf{a}_i (1 \leq i \leq R)$, and thus, it can model the self-scattering effects; and 2) it does not force the assumption that $\zeta_{ij} = \gamma_{ij} x_i x_j (1 \leq i, j \leq R)$ as in GBM and PPNMM, and thus, it can handle the situations when the reflections from a neighboring pixel contribute to the bilinear terms but not to the linear abundance term of the target pixel.

III. Sparse Abundance Estimation Method

A. Sparse Linear Regression via SMV-ADMM

Recent research in [28] and [29] shows that the abundance estimation problem can be formulated as a linear sparse regression problem and could be efficiently solved under the sparse representation [43] framework. Specifically, with a given dictionary $\mathbf{A}$ and a mixed pixel $\mathbf{y}$, we can formulate the problem as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, s.t. \|\mathbf{y} - \mathbf{Ax}\|_2 \leq \epsilon_1$$
$$\mathbf{x} \geq \mathbf{0}, \mathbf{1}^T\mathbf{x} = 1 \tag{8}$$

where $\|\cdot\|_0$ denotes the l_0-norm, which is defined as the number of nonzero entries in the vector of interest, and $\epsilon_1 \geq 0$ is the error tolerance level due to noises and various modeling errors. This problem is nonconvex and generally NP-hard to solve. Instead, if the solution is sufficiently sparse, it can be relaxed to a linear programming problem by replacing the l_0-norm with the l_1-norm. However, because the abundance vector is constrained by both ASC and ANC, its l_1-norm remains as a constant equaling to unity, rendering the whole l_1-minimization problem meaningless. Therefore, in [28] and [29], they propose to relax ASC and solve an equivalent *lasso* problem [44] with only the nonnegative constraint

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda_1\|\mathbf{x}\|_1, s.t. \mathbf{x} \geq \mathbf{0} \tag{9}$$

where $\|\mathbf{x}\|_1 = \sum_{i=1}^{R}|x_i|$ and $\lambda_1 > 0$ is actually a Lagrangian multiplier with $\lambda_1 \to 0$ as $\epsilon_1 \to 0$. This optimization problem is convex and can be solved by the current l_1-minimization techniques such as the nonnegative constraint split Bregman method [45] and SMV-ADMM algorithm [46] (named CSUnSAL+ in [29]). The differences between the two approaches are as follows.

1) In [28], they tried to solve an overconstrained l_1-regression problem with a "tall and thin" dictionary $\mathbf{A}(s.t.\ L > R)$, which is also known as a least absolute deviations problem. The dictionary $\mathbf{A}$ consists of the pure endmembers detected by the N-Finder algorithm [10]. The solution of this problem is known to be robust and less sensitive to the presence of noise and outliers [47].
2) Iordache *et al.* [29] proposed to solve a sparse recovery problem with a "short and fat" dictionary $\mathbf{A}(s.t.\ L < R)$ under the l_1 constraint. They propose to use a huge dictionary $\mathbf{A}$, which consists of selected pure endmembers in the USGS library. From the compressed sensing theory [43], to guarantee the uniqueness of the sparsely recovered solution for an underdetermined system, the matrix $\mathbf{A}$ is required to have low mutual coherence $\mu_{co}(\mathbf{A})$, which is defined as

$$\mu_{co}(\mathbf{A}) = \max_{1\leq i,j\leq R, i\neq j} \frac{\left|\mathbf{a}_i^T\mathbf{a}_j\right|}{\|\mathbf{a}_i\|_2\|\mathbf{a}_j\|_2}. \tag{10}$$

However, in practice, the pure endmembers tend to be similar and highly correlated, rendering the coherence of the dictionary $\mu_{co}(\mathbf{A})$ to be large. As it has been shown in [29], even with the pruned USGS dictionary, the mutual coherence of the matrix $\mu_{co}(\mathbf{A})$ nearly equals to unity (i.e., $\mu_{co}(\mathbf{A}) = 0.9986$). In such cases, even though the abundance vector may be very sparse, the uniqueness of the solution cannot be guaranteed. Hence, the unmixing result might not be accurate.

Therefore, in our proposed method, we solve an overconstrained l_1-regression problem, where the endmember dictionary $\mathbf{A}$ is made up of pure endmembers extracted by the VCA [6] from the hyperspectral image itself.

To solve the problem (9), we introduce and adopt the nonnegative SMV-ADMM algorithm due to its efficiency and robustness. First, we introduce an auxiliary variable $\mathbf{z}$ and transform the problem as follows:

$$\min_{\mathbf{x},\mathbf{z}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda_1\|\mathbf{z}\|_1$$
$$s.t.\ \mathbf{x}-\mathbf{z} = \mathbf{0}, \mathbf{z} \geq \mathbf{0} \tag{11}$$

then, the corresponding augmented Lagrangian function w.r.t. $\mathbf{x}$ and $\mathbf{z}$ can be formed as

$$L_{\mu_1}(\mathbf{x},\mathbf{z},\boldsymbol{t}) = \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda_1\|\mathbf{z}\|_1 + <\mathbf{t}, \mathbf{x}-\mathbf{z}> + \frac{\mu_1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 \tag{12}$$

where $\mathbf{t}$ is the Lagrangian multiplier for the equality constraint $\mathbf{x}-\mathbf{z}=\mathbf{0}$ and $\mu_1 > 0$ is a penalty parameter. Function (12) is constrained by $\mathbf{z} \geq \mathbf{0}$; it can be minimized w.r.t. $\mathbf{x}$ and $\mathbf{z}$ iteratively by fixing one of the variables and updating the other. The entire algorithm is summarized in Algorithm 1, where we replace the variable $\mathbf{t}$ by $\boldsymbol{\tau} = \mathbf{t}/\mu_1$ to simplify all the equations.

Algorithm 1 Nonnegative SMV-ADMM

Input: The observation $\mathbf{y}$, the endmember dictionary $\mathbf{A}$, and the balancing parameter λ_1;
Output: The estimated abundance $\hat{\mathbf{x}}$;
1: Initialize: $\mathbf{x}^0$, $\mathbf{z}^0$, $\boldsymbol{\tau}^0$, μ_1, $k=0$;
2: **while** not converged **do**
3: Fix $\mathbf{z}$ and update $\mathbf{x}$ by:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} L_{\mu_1}(\mathbf{x}, \mathbf{z}^k, \boldsymbol{\tau}^k)$$
$$= (\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{I})^{-1}\left(\mathbf{A}^T\mathbf{y} + \mu_1(\mathbf{z}^k - \boldsymbol{\tau}^k)\right)$$

4: Fix $\mathbf{x}$ and update $\mathbf{z}$ by:

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}\geq\mathbf{0}} L_{\mu_1}(\mathbf{x}^{k+1}, \mathbf{z}, \boldsymbol{\tau}^k)$$
$$= \arg\min_{\mathbf{z}\geq\mathbf{0}} \lambda_1\|\mathbf{z}\|_1 + \frac{\mu_1}{2}\left\|\mathbf{z} - (\mathbf{x}^{k+1} + \boldsymbol{\tau}^k)\right\|_2^2$$
$$= \max\left[S_{\lambda_1/\mu_1}(\mathbf{x}^{k+1} + \boldsymbol{\tau}^k), 0\right]$$

5: Update the Lagrangian multiplier $\boldsymbol{\tau}$:

$$\boldsymbol{\tau}^{k+1} = \boldsymbol{\tau}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}$$

6: Update $k: k = k+1$.
7: **end while**
8: **return** $\hat{\mathbf{x}} = \mathbf{z}^k$.

In Step 3, note that, with $\mu_1 > 0$, $\mathbf{A}^T\mathbf{A} + \mu_1\mathbf{I}$ is always invertible, so updating $\mathbf{x}$ is essentially a ridge regression problem. To update the variable $\mathbf{z}$, the subproblem in Step 4 is solved by a soft-shrinkage thresholding operator $S_{\kappa_1}(\cdot)$ [48] introduced in the following lemma.

Lemma 1: Consider the following optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \kappa_1\|\boldsymbol{\theta}\|_1 + \frac{1}{2}\|\boldsymbol{\theta} - \mathbf{c}\|_2^2 \tag{13}$$

where $\boldsymbol{\theta}^* \in \mathbb{R}^n$ is the optimal solution vector, $\mathbf{c}$ is a constant vector of the same size as $\boldsymbol{\theta}^*$, and $\kappa_1 > 0$ is a penalty parameter. Then, the ith component θ_i^* of vector $\boldsymbol{\theta}^*$ should be

$$\theta_i^* = S_{\kappa_1}(c_i) = \begin{cases} c_i - \kappa_1, & \text{if } c_i > \kappa_1 \\ 0, & \text{if } |c_i| \le \kappa_1 \\ c_i + \kappa_1, & \text{if } c_i < -\kappa_1 \end{cases} \tag{14}$$

where c_i is the ith component of $\mathbf{c}$ and $S_{\kappa_1}(\cdot)$ is a soft-shrinkage thresholding operator with threshold κ_1.

Remark 1: Stopping criteria: In the ADMM algorithm, we usually use the primal residue r_p and dual residue r_d as the indicators of convergence

$$r_p^k = \|\mathbf{x}^k - \mathbf{z}^k\|_2$$
$$r_d^k = \mu_1 \|\mathbf{z}^{k+1} - \mathbf{z}^k\|_2.$$

The algorithm stops when both the residues are smaller than certain preset tolerances. Otherwise, the algorithm stops when the maximum iteration is reached.

B. Extension to Nonlinear Abundance Estimation Problem

1) Problem Transformation Under ANC: By carefully examining (7), it is interesting to note that MGBM can be seen as the LMM with R original endmembers and $R^* = (1/2)R(R+1)$ correlated endmembers. More specifically, by considering each second-order spectral term $\mathbf{a}_i \odot \mathbf{a}_j (1 \le i \le j \le R)$ as a new spectral component associated with the fraction ζ_{ij}, the model can be rewritten as a linear combination of all spectra

$$\mathbf{y} = \sum_{k=1}^{R} x_k \mathbf{a}_k + \sum_{l=1}^{R^*} e_l \mathbf{b}_l + \mathbf{n} \tag{15}$$

where

$$e_l = \zeta_{ij}, \mathbf{b}_l = \mathbf{a}_i \odot \mathbf{a}_j$$
$$l = j + \frac{(2R-i)(i-1)}{2}, \qquad 1 \le i \le j \le R. \tag{16}$$

If we define the bilinear dictionary and its corresponding bilinear representation as $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{R^*}]$ and $\mathbf{e} = [e_1, e_2, \ldots, e_{R*}]^T$, respectively, then we can rewrite (15) as

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} + \mathbf{n} \\ &= [\mathbf{A}, \mathbf{B}] \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} + \mathbf{n} \\ &= \mathbf{M}\boldsymbol{\phi} + \mathbf{n}. \end{aligned} \tag{17}$$

Here, $\mathbf{M} = [\mathbf{A}, \mathbf{B}]$ is a composite dictionary and $\boldsymbol{\phi} = [\mathbf{x}^T, \mathbf{e}^T]^T$ is the corresponding composite representation. Therefore, the bilinear abundance estimation problem can be transformed and solved in the sparse linear regression framework as

$$\min_{\boldsymbol{\phi}} \frac{1}{2}\|\mathbf{y} - \mathbf{M}\boldsymbol{\phi}\|_2^2 + \lambda_1' \|\boldsymbol{\phi}\|_1, s.t.\ \boldsymbol{\phi} \ge \mathbf{0} \tag{18}$$

where $\lambda_1' > 0$. This new sparse regression problem can still be efficiently solved by the nonnegative constraint SMV-ADMM algorithm in the same manner. Once the recovered sparse code $\hat{\boldsymbol{\phi}} = [\hat{\mathbf{x}}^T, \hat{\mathbf{e}}^T]^T$ is obtained, the pure endmembers will be identified, and their associated abundances can be estimated from the vector $\hat{\mathbf{x}}$. Hence, we can effectively get rid of the small annoying bilinear components in MGBM and accurately identify and predict the abundances for the linear combinations.

2) Enforcing the ASC: Compared with the original problem in (9) for LMM, the extended problem (18) for BMM minimizes the l_1-norm of the composite sparse representation $\boldsymbol{\phi}$ rather than $\|\mathbf{x}\|_1$. Although the ASC in (9) could not be enforced as explained in Section III-A, in (18), the ASC could be partially enforced on the abundance vector $\mathbf{x}$ instead of the entire composite abundance $\boldsymbol{\phi}$. Therefore, we could add the ASC to (18) as

$$\min_{\boldsymbol{\phi}} \frac{1}{2}\|\mathbf{y} - \mathbf{M}\boldsymbol{\phi}\|_2^2 + \lambda_1' \|\boldsymbol{\phi}\|_1$$
$$s.t.\ \boldsymbol{\phi} \ge \mathbf{0}, \mathbf{k}^T\boldsymbol{\phi} = 1 \tag{19}$$

where $\mathbf{k} = [\mathbf{1}^T, \mathbf{0}^T]^T$ ($\mathbf{1} \in \mathbb{R}^R$, and $\mathbf{0} \in \mathbb{R}^{R^*}$). Instead of directly solving the equality constraint problem in (19), we solve the following relaxed problem instead:

$$\min_{\boldsymbol{\phi} \ge 0} \frac{1}{2} \left\| \begin{bmatrix} \mathbf{y} \\ \delta_1 \end{bmatrix} - \begin{bmatrix} \mathbf{M} \\ \delta_1 \mathbf{k}^T \end{bmatrix} \boldsymbol{\phi} \right\|_2^2 + \lambda_1' \|\boldsymbol{\phi}\|_1 \tag{20}$$

where $\delta_1 > 0$ is a regularization parameter and (20) is still a nonnegative constraint sparse regression problem that can be efficiently solved by the nonnegative constraint SMV-ADMM algorithm. Therefore, the ASC, which could not be enforced under the LMM [28], [29], now can be incorporated in our proposed new framework. In Section VI-C, we demonstrate on the synthetic data that the abundance estimation performance could be further improved with the ASC and qualitatively analyze the choice of the parameter δ_1. Some theoretical analysis for choosing the optimal δ_1 remains for the future work.

For the real hyperspectral images, the effects of ASC need to be further studied. Sometimes, the ASC is prone to some criticisms for being a too strict constraint [28], [29]. Without the ground truth of the abundance information for the real data, the validation of the ASC still remains as a problem.

IV. Joint-Sparsity Model for Abundance Estimation

Following the notations in the previous section, now let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{L \times N}$ be an observation matrix, which contains N bilinear mixed pixels within a sliding window of fixed size, and let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{R \times N}$ and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N] \in \mathbb{R}^{R^* \times N}$ be the linear and bilinear abundance representation matrices associated with the dictionaries $\mathbf{A}$ and $\mathbf{B}$, respectively.

A. Problem Formulation

Expecting adjacent pixels in a small sliding window to have very similar endmembers, we enforce a joint-sparse representation on the abundance matrix $\boldsymbol{\Phi} = [\mathbf{X}^T, \mathbf{E}^T]^T = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N] \in \mathbb{R}^{(R+R^*) \times N}$ which contains the target pixel and its neighbors, where all abundance vectors $\boldsymbol{\phi}_i (1 \le i \le N)$ should have the same support set and satisfy $\mathbf{Y} = \mathbf{M}\boldsymbol{\Phi}$.

We slide the window through the whole image one pixel at a time to estimate the abundance for every pixel in the center of the window. To seek a row-sparse solution for each $\mathbf{\Phi}$, we need to minimize the number of nonzero rows in matrix $\mathbf{\Phi}$. Thus, the problem can be formulated as follows:

$$\min_{\mathbf{\Phi}} \|\mathbf{\Phi}\|_{0,2}, s.t. \|\mathbf{Y} - \mathbf{M}\mathbf{\Phi}\|_F \leq \epsilon_2$$
$$\mathbf{\Phi} \geq \mathbf{0}, \mathbf{k}^T\mathbf{\Phi} = \mathbf{1}^T \tag{21}$$

where $\|\mathbf{\Phi}\|_{0,2}$ provides a measure for the number of "active" rows in $\mathbf{\Phi}$ that contain nonzero entries, $\|\cdot\|_F$ is the Frobenius norm, $\epsilon_2 > 0$ is the noise tolerance, and $\mathbf{1} \in \mathbb{R}^N$. Similar to the l_0-minimization problem in (7), this problem is also a combinatorial optimization problem and thus NP-hard in general. The same as the case in (19), the matrix $l_{0,2}$-norm in (21) can be replaced by the matrix $l_{1,2}$-norm. Thus, the problem can be reformulated as

$$\min_{\mathbf{\Phi}\geq\mathbf{0}} \frac{1}{2}\left\|\begin{bmatrix}\mathbf{Y}\\ \delta_2\mathbf{1}^T\end{bmatrix} - \begin{bmatrix}\mathbf{M}\\ \delta_2\mathbf{k}^T\end{bmatrix}\mathbf{\Phi}\right\|_F^2 + \lambda_2\|\mathbf{\Phi}\|_{1,2} \tag{22}$$

where $\delta_2 > 0$ and $\lambda_2 > 0$ are the regularization parameters, $\|\mathbf{\Phi}\|_{1,2} = \sum_{i=1}^{R+R^*}\|\boldsymbol{\phi}^i\|_2$, and $\boldsymbol{\phi}^i \in \mathbb{R}^N$ is the ith row of the matrix $\mathbf{\Phi}$. This problem (22) now is convex and can be efficiently solved by the nonnegative constraint MMV-ADMM algorithm proposed in the next section. Once the composite sparse representation $\hat{\mathbf{\Phi}} = [\hat{\mathbf{X}}^T, \hat{\mathbf{E}}^T]^T$ is derived, the abundances for each pixel can be obtained from $\hat{\mathbf{X}}$.

Remark 2: Discussion on the limitation of the joint-sparse constraint: The proposed joint-sparsity assumption forces all the pixels in the sliding window to share the same support set. The joint-sparsity constraint is effective for unmixing homogeneous regions, where most of the pixels consist of very similar types of endmembers. However, when the sliding window contains subpixel targets or noncontiguous materials, such a strict spatial constraint might limit the ability to accurately estimate the proportions of endmembers (see Fig. 2).

B. Recovery of Jointly Sparse Abundance Vectors via MMV-ADMM Algorithm

In this section, we introduce the MMV-ADMM algorithm to solve the following problem:

$$\min_{\mathbf{\Phi}\geq\mathbf{0}} \frac{1}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{\Phi}\|_F + \lambda_2\|\mathbf{\Phi}\|_{1,2}. \tag{23}$$

It should be noticed that the problem (22) is exactly in the form of (23) if we let $\hat{\mathbf{Y}} = \begin{bmatrix}\mathbf{Y}\\ \delta_2\mathbf{1}^T\end{bmatrix}$ and $\hat{\mathbf{M}} = \begin{bmatrix}\mathbf{M}\\ \delta_2\mathbf{k}^T\end{bmatrix}$ and reformulate the objective function as $(1/2)\|\hat{\mathbf{Y}} - \hat{\mathbf{M}}\mathbf{\Phi}\|_F + \lambda_2\|\mathbf{\Phi}\|_{1,2}$.

Similar to the SMV-ADMM algorithm, we first introduce an auxiliary matrix variable $\mathbf{Z} \in \mathbb{R}^{(R+R^*)\times N}$ and transform the problem into

$$\min_{\mathbf{\Phi},\mathbf{Z}} \frac{1}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{\Phi}\|_F^2 + \lambda_2\|\mathbf{Z}\|_{1,2}$$
$$s.t.\ \boldsymbol{Z} \geq \mathbf{0}, \mathbf{\Phi} - \mathbf{Z} = \mathbf{0}. \tag{24}$$

Thus, the augmented Lagrangian function of (24) w.r.t. $\mathbf{\Phi}$ and $\mathbf{Z}$ can be formed as

$$L_{\mu_2}(\mathbf{\Phi}, \mathbf{Z}, \mathbf{T}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{\Phi}\|_F^2 + \lambda_2\|\mathbf{Z}\|_{1,2} + < \mathbf{T}, \mathbf{\Phi} - \mathbf{Z} > + \frac{\mu_2}{2}\|\mathbf{\Phi} - \mathbf{Z}\|_F^2 \tag{25}$$

where $\mu_2 > 0$ is a penalty parameter, $\mathbf{T} \in \mathbb{R}^{(R+R^*)\times N}$ is a matrix of Lagrangian multipliers, and we let $\mathbf{\Lambda} = \mathbf{T}/\mu_2$. Similarly, we minimize the augmented Lagrangian function iteratively by fixing one variable and updating the other. The complete algorithm is summarized in Algorithm 2.

Algorithm 2 Nonnegative MMV-ADMM

Input: The pixel array $\{\mathrm{Y}\}$, the composite dictionary $\mathbf{\Phi}$, and the balancing parameter λ_2;
Output: The estimated abundance $\hat{\mathbf{X}}$;
1: Initialize: $\mathbf{\Phi}^0, \mathbf{Z}^0, \mathbf{\Lambda}^0, \mu_2, k = 0$;
2: **while** not converged **do**
3: Fix $\mathbf{Z}$ and update $\mathbf{\Phi}$ by:

$$\mathbf{\Phi}^{k+1} = \arg\min_{\mathbf{\Phi}} L_{\mu_2}(\mathbf{\Phi}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$$
$$= (\mathbf{M}^T\mathbf{M} + \mu_2\mathbf{I})^{-1}\left(\mathbf{M}^T\mathbf{Y} + \mu_2(\mathbf{Z}^k - \mathbf{\Lambda}^k)\right)$$

4: Fix $\mathbf{\Phi}$ and update $\mathbf{Z}$ by:

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}\geq\mathbf{0}} L_{\mu_2}(\mathbf{\Phi}^{k+1}, \mathbf{Z}, \mathbf{\Lambda}^k)$$
$$= \arg\min_{\mathbf{Z}\geq\mathbf{0}} \lambda_2\|\mathbf{Z}\|_{2,1} + \frac{\mu_2}{2}\left\|\mathbf{Z} - (\mathbf{\Phi}^{k+1} + \mathbf{\Lambda}^k)\right\|_2^2$$
$$= \max\left[S^*_{\lambda_2/\mu_2}(\mathbf{\Phi}^{k+1} + \mathbf{\Lambda}^k), 0\right]$$

5: Update the Lagrangian multiplier $\mathbf{\Lambda}$:

$$\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k + \mathbf{\Phi}^{k+1} - \mathbf{Z}^{k+1}$$

6: Update k: $k = k + 1$.
7: **end while**
8: **return** $[\hat{\mathbf{X}}^T, \hat{\mathbf{E}}^T]^T = {}^k$.

Compared with the SMV-ADMM algorithm, here, we try to recover a row-sparse matrix jointly rather than reconstruct each sparse vector individually; therefore, the $\mathbf{Z}$ updating in Step 4 is solved by the row-shrinkage thresholding operator $S^*_{\kappa_2}(\cdot)$ [49] introduced in the following lemma instead.

Lemma 2: Consider the following optimization problem:

$$\mathbf{\Theta}^* = \arg\min_{\mathbf{\Theta}} \kappa_2\|\mathbf{\Theta}\|_{1,2} + \frac{1}{2}\|\mathbf{\Theta} - \mathbf{C}\|_F^2 \tag{26}$$

where $\mathbf{\Theta}^* \in \mathbb{R}^{m\times n}$ is the optimal solution matrix, $\mathbf{C}$ is a constant matrix of the same size as $\mathbf{\Theta}^*$, and $\kappa_2 > 0$ is a penalty parameter. The minimizer of (26) is given by

$$\mathbf{\Theta}^* = S^*_{\kappa_2}(\mathbf{C})$$

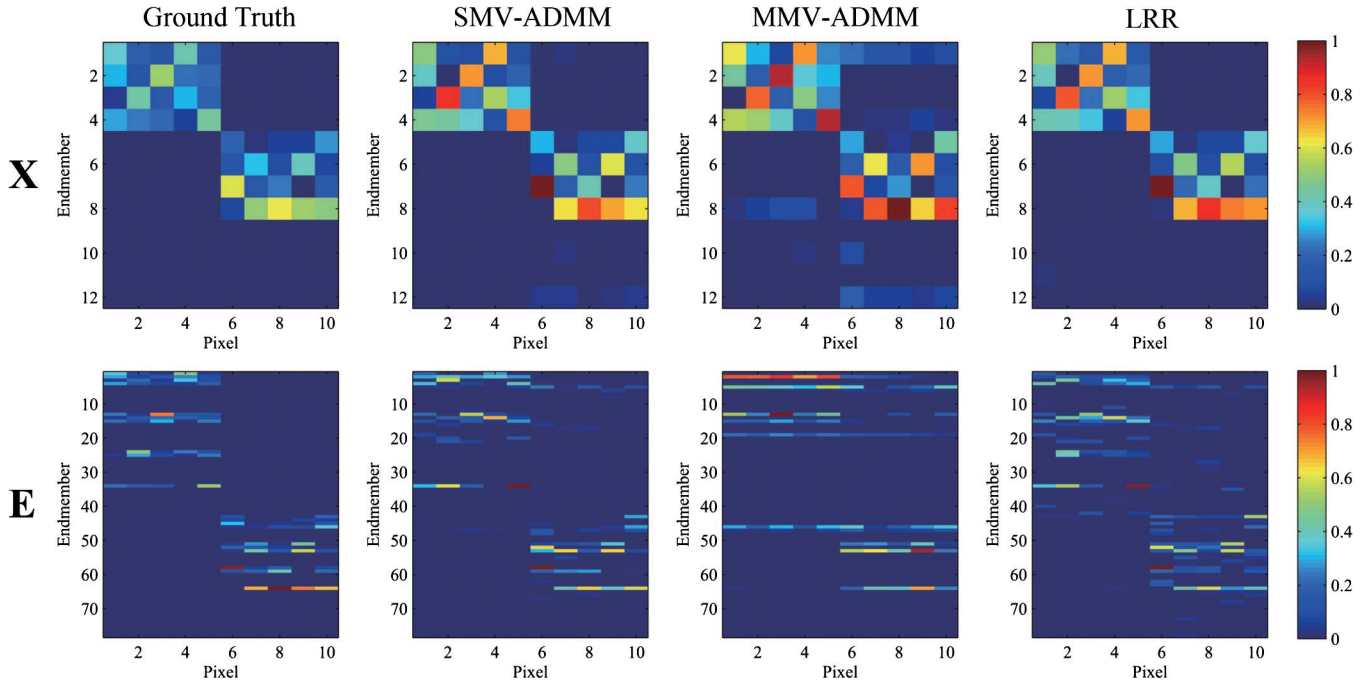


Fig. 2. Results of the estimated abundance matrix $\mathbf{X}$ and bilinear coefficient matrix $\mathbf{E}$ for the SMV-ADMM, MMV-ADMM, and LRR methods. The first row shows the ground-truth and estimated abundance matrices, and the second row shows the ground-truth and estimated bilinear coefficient matrices. From left to right are the original data and the results from SMV-ADMM, MMV-ADMM, and LRR, respectively.

where the function $S^*_{\kappa_2}(\cdot)$ is defined as follows: Let the ith rows of matrices $\mathbf{\Theta}^*$ and $\mathbf{C}$ be $\boldsymbol{\theta}^{i^*}$ and $\mathbf{c}^i$, respectively; then

$$\boldsymbol{\theta}^{i^*} = S^*_{\kappa_2}(\mathbf{c}^i) = \begin{cases} \frac{\|\mathbf{c}^i\|_2 - \kappa_2}{\|\mathbf{c}^i\|_2}\mathbf{c}^i, & \text{if } \kappa_2 < \|\mathbf{c}^i\|_2 \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

V. LRR Model for Abundance Estimation

A. Problem Formulation

In the previous section, the joint-sparsity model is introduced for the abundance estimation problem, which exploits the structured sparsity of the abundances between neighboring pixels. However, as has already been analyzed in Remark 2, the joint-sparsity model introduces some *aliasing artifacts* on the boundaries when the neighboring pixels consist of different endmembers. Therefore, instead of using the row-sparsity property of the abundance matrix, we further propose to exploit its rank property by the following theorem [50].

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

$$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{Y}) = k. \tag{28}$$

According to the aforementioned theorem, our dictionary $\mathbf{A}$ usually satisfies the full column rank property as the extracted pure endmembers are generally distinct from each other and the dimension of hyperspectral data L is larger than the total number of endmembers R. If the columns of $\mathbf{Y}$ are highly correlated, which means that the matrix $\mathbf{Y}$ is a low-rank matrix, it further indicates that the corresponding representation matrix $\mathbf{X}$ is also low rank.

To use this property, we employ the recently proposed LRR model [39], [40] for our joint abundance estimation problem. LRR seeks the lowest rank representation of the abundance matrix to capture the spatial structure of the data jointly. From Fig. 2, we can see that the proposed LRR algorithm can alleviate the aliasing problem caused by the strict row-sparsity regularization. Next, we introduce our LRR algorithm for solving the joint abundance estimation problem.

Starting with the simple LMM, for the given data $\mathbf{Y}$, we seek a low-rank abundance representation $\mathbf{X}$ with ANC by solving the following optimization problem:

$$\begin{aligned} &\mathbf{X}^* = \arg\min_{\mathbf{X}} \text{rank}(\mathbf{X}), s.t.\ \mathbf{X} \geq \mathbf{0} \\ &\mathbf{Y} - \mathbf{AX} = \mathbf{0}, \mathbf{X} \geq \mathbf{0}, \mathbf{1}^T\mathbf{X} = \mathbf{1}^T \end{aligned} \tag{29}$$

where $\mathbf{1} \in \mathbb{R}^N$ and $\mathbf{X}^*$ is the lowest rank solution with the ASC and ANC constraints.

For the abundance estimation problem of MGBM, we factorize the bilinear terms into a bilinear dictionary $\mathbf{B} \in \mathbb{R}^{L\times R^*}$ and its associated sparse bilinear abundance matrix $\mathbf{E} \in \mathbb{R}^{R^*\times N}$. Thus, we propose to solve a dual representation problem for MGBM as

$$\begin{aligned} &\min_{\mathbf{X}\geq\mathbf{0},\mathbf{E}\geq\mathbf{0}} \text{rank}(\mathbf{X}) + \lambda_3\|\mathbf{E}\|_0 \\ &s.t.\ \mathbf{Y} - \mathbf{AX} - \mathbf{BE} = \mathbf{0}, \mathbf{1}^T\mathbf{X} = \mathbf{1}^T \end{aligned} \tag{30}$$

where $\lambda_3 > 0$ is a parameter used to balance the effects of the two terms.

If we choose $r < R$ endmembers for MGBM, then the number of bilinear coefficients for one pixel that are nonzero is $r^* = (r(r+1))/(2) \ll R^*$. Therefore, $\mathbf{E}$ is definitely a sparse matrix. We do not enforce the low-rank regularization on $\mathbf{E}$ because $\mathbf{E}$ is much sparser than $\mathbf{X}$ and LRR will produce "denser" results than the l_0-norm constraint [51]. Here, we have chosen to enforce the l_0-norm regularization on $\mathbf{E}$ instead of matrix $l_{0,2}$-norm (joint-sparsity constraint) for two reasons: 1) We have analyzed the performance of the joint-sparsity model on the synthetic data, and it causes some *aliasing artifacts* on the boundaries between two inhomogeneous regions; and 2) the computational cost for joint sparsity is higher, and simulation results on real HSI data show little difference between the two norms.

However, the aforementioned optimization problem (30) is still difficult to solve due to the discrete nature of the rank function and the nonconvexity of the l_0-norm. Fortunately, according to our previous discussion in Section III, the l_0-norm can be replaced by the l_1-norm, and as suggested by the matrix completion methods [52], the nuclear norm [53] provides a good surrogate for the matrix rank function. Hence, instead of (30), we seek the solution of an alternative problem as follows:

$$\min_{\mathbf{X}\geq\mathbf{0},\mathbf{E}\geq\mathbf{0}} \|\mathbf{X}\|_* + \lambda_3\|\mathbf{E}\|_1$$
$$s.t.\ \mathbf{Y} - \mathbf{AX} - \mathbf{BE} = \mathbf{0}, \mathbf{1}^T\mathbf{X} = \mathbf{1}^T \tag{31}$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix defined as

$$\|\mathbf{X}\|_* = \text{trace}(\sqrt{\mathbf{X}^T\mathbf{X}}) = \sum_{i=1}^{\min\{R,N\}} \sigma_i \tag{32}$$

here, $\sigma_i (0 \leq i \leq \min\{R, N\})$ are singular values of the abundance matrix $\mathbf{X}$. To handle ASC, we solve the following problem instead:

$$\min_{\mathbf{X}\geq\mathbf{0},\mathbf{E}\geq\mathbf{0}} \|\mathbf{X}\|_* + \lambda_3\|\mathbf{E}\|_1$$
$$s.t. \begin{bmatrix}\mathbf{Y}\\ \delta_3\mathbf{1}^T\end{bmatrix} - \begin{bmatrix}\mathbf{A}\\ \delta_3\mathbf{1}^T\end{bmatrix}\mathbf{X} - \begin{bmatrix}\mathbf{B}\\ \mathbf{0}^T\end{bmatrix}\mathbf{E} = \mathbf{0} \tag{33}$$

where $\delta_3 > 0$. The problem stated in (33) is convex, and it can be efficiently solved by state-of-the-art convex optimization techniques. Moreover, a close examination of (31) reveals that, if we let $\mathbf{B} = \mathbf{I}$, then the problem degenerates to the original LRR problem introduced in [40] under the constraints of ASC and ANC. Hence, our problem is a more general case of the original LRR model.

B. Low-Rank Abundance Estimation by Convex Optimization

In this section, we try to solve the following problem:

$$\min_{\mathbf{X}\geq\mathbf{0},\mathbf{E}\geq\mathbf{0}} \|\mathbf{X}\|_* + \lambda_3\|\mathbf{E}\|_1$$
$$s.t.\ \mathbf{Y} - \mathbf{AX} - \mathbf{BE} = \mathbf{0}. \tag{34}$$

Similar to our discussion for the nonnegative MMV-ADMM algorithm, the original problem in (33) is in the same form as (34). In [40], the original LRR problem is solved by inexact augmented Lagrangian method (IALM) [54]. However, because the sparse error matrix in the original LRR model has to be factorized into a bilinear dictionary $\mathbf{B}$ and an associated bilinear representation $\mathbf{E}$ in (34), the original optimization algorithm needs to be modified. By adding two auxiliary matrices $\mathbf{P} \in \mathbb{R}^{R\times N}$ and $\mathbf{Q} \in \mathbb{R}^{R^*\times N}$, the problem in (34) can be reformulated as

$$\min_{\mathbf{P}\geq\mathbf{0},\mathbf{Q}\geq\mathbf{0}} \|\mathbf{P}\|_* + \lambda_3\|\mathbf{Q}\|_1, s.t\ \mathbf{X} - \mathbf{P} = \mathbf{0}$$
$$\mathbf{Q} - \mathbf{E} = \mathbf{0}, \mathbf{Y} - \mathbf{AX} - \mathbf{BE} = \mathbf{0}. \tag{35}$$

Thus, the augmented Lagrangian function w.r.t. $\mathbf{X}$, $\mathbf{E}$, $\mathbf{P}$, and $\mathbf{Q}$ can be formed as

$$\begin{aligned} L_{\mu_3}(\mathbf{X},\mathbf{E},\mathbf{P},\mathbf{Q}) &= \|\mathbf{P}\|_* + \lambda_3\|\mathbf{Q}\|_1 \\ &\quad + \text{tr}\left[\mathbf{T}_1^T(\mathbf{Y} - \mathbf{AX} - \mathbf{BE})\right] + \text{tr}\left[\mathbf{T}_2^T(\mathbf{X} - \mathbf{P})\right] \\ &\quad + \text{tr}\left[\mathbf{T}_3^T(\mathbf{E} - \mathbf{Q})\right] \\ &\quad + \frac{\mu_3}{2}(\|\mathbf{Y} - \mathbf{AX} - \mathbf{BE}\|_F^2 + \|\mathbf{P} - \mathbf{X}\|_F^2 + \|\mathbf{Q} - \mathbf{E}\|_F^2) \end{aligned} \tag{36}$$

where $\mu_3 > 0$ is a penalty parameter; $\mathbf{T}_1 \in \mathbb{R}^{L\times N}$, $\mathbf{T}_2 \in \mathbb{R}^{R\times N}$, and $\mathbf{T}_3 \in \mathbb{R}^{R^*\times N}$ are the matrices of Lagrangian multipliers. To simplify the equations, we let $\mathbf{\Lambda}_1 = \mathbf{T}_1/\mu_3$, $\mathbf{\Lambda}_2 = \mathbf{T}_2/\mu_3$, and $\mathbf{\Lambda}_3 = \mathbf{T}_3/\mu_3$. Following the same optimization scheme in the previous sections, the modified nonnegative IALM optimization algorithm for our problem is summarized in Algorithm 3.

Algorithm 3 Nonnegative Constraint IALM for Low-Rank Representation Recovery

Input: The observation matrix $\mathbf{Y}$, the linear dictionary $\mathbf{A}$, the bilinear dictionary $\mathbf{B}$, the scalar $\rho = 1.1$, and the balancing parameter λ_3;

Output: The estimated abundance $\hat{\mathbf{X}}$ and bilinear representation matrix $\hat{\mathbf{E}}$;

1: Initialize: $\mathbf{X}^0, \mathbf{P}^0, \mathbf{E}^0, \mathbf{Q}^0, \mathbf{\Lambda}_1^0, \mathbf{\Lambda}_2^0, \mathbf{\Lambda}_3^0, \mu_3, \mu_{\max} = 10^6$, $k = 0$;

2: **while** not converged **do**

3: Fix $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{E}$ and update $\mathbf{P}$ by:

$$\begin{aligned} \mathbf{P}^{k+1} &= \arg\min_{\mathbf{P}\geq\mathbf{0}} L_{\mu_3}(\mathbf{X}^k,\mathbf{E}^k,\mathbf{P},\mathbf{Q}^k) \\ &= \arg\min_{\mathbf{P}\geq\mathbf{0}} \|\mathbf{P}\|_* + \frac{\mu_3}{2}\left\|\mathbf{P} - \left(\mathbf{X}^k + \mathbf{\Lambda}_2^k\right)\right\|_F^2 \\ &= \max\left[\mathcal{D}_{1/\mu_3}\left(\mathbf{X}^k + \mathbf{\Lambda}_2^k\right), 0\right] \end{aligned}$$

4: Fix $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{E}$ and update $\mathbf{X}$ by:

$$\begin{aligned} \mathbf{X}^{k+1} &= \arg\min_{\mathbf{J}} L_{\mu_3}(\mathbf{X},\mathbf{E}^k,\mathbf{P}^{k+1},\mathbf{Q}^k) \\ &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left[\mathbf{A}^T(\mathbf{Y} - \mathbf{BE}^k) + \mathbf{P}^{k+1}\right. \\ &\qquad \left. + \mathbf{A}^T\mathbf{\Lambda}_1^k - \mathbf{\Lambda}_2^k\right] \end{aligned}$$

5: Fix $\mathbf{X}$, $\mathbf{P}$, $\mathbf{E}$ and update $\mathbf{Q}$ by:

$$\begin{aligned}\mathbf{Q}^{k+1} &= \arg\min_{\mathbf{Q}\geq\mathbf{0}} L_{\mu_3}(\mathbf{X}^{k+1}, \mathbf{E}^k, \mathbf{P}^{k+1}, \mathbf{Q}) \\ &= \arg\min_{\mathbf{Q}\geq\mathbf{0}} \lambda_3\|\mathbf{Q}\|_1 + \frac{\mu_3}{2}\left\|\mathbf{Q} - \left(\mathbf{E}^k + \mathbf{\Lambda}_3^k\right)\right\|_F^2 \\ &= \max\left[S_{\lambda_3/\mu_3}\left(\mathbf{E}^k + \mathbf{\Lambda}_3^k\right), 0\right]\end{aligned}$$

6: Fix $\mathbf{X}$, $\mathbf{P}$, $\mathbf{Q}$ and update $\mathbf{E}$ by:

$$\begin{aligned}\mathbf{E}^{k+1} &= \arg\min_{\mathbf{E}} L_{\mu_3}(\mathbf{X}^{k+1}, \mathbf{E}, \mathbf{P}^{k+1}, \mathbf{Q}^{k+1}) \\ &= (\mathbf{I} + \mathbf{B}^T\mathbf{B})^{-1}\left[\mathbf{B}^T(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k+1}) + \mathbf{Q}^{k+1}\right. \\ &\qquad \left. + \mathbf{B}^T\mathbf{\Lambda}_1^k - \mathbf{\Lambda}_3^k\right]\end{aligned}$$

7: Update Lagrange multiplier $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, $\mathbf{\Lambda}_3$:

$$\begin{aligned}\mathbf{\Lambda}_1^{k+1} &= \mathbf{\Lambda}_1^k + \mathbf{Y} - \mathbf{A}\mathbf{X}^{k+1} - \mathbf{B}\mathbf{E}^{k+1} \\ \mathbf{\Lambda}_2^{k+1} &= \mathbf{\Lambda}_2^k + \mathbf{X}^{k+1} - \mathbf{P}^{k+1} \\ \mathbf{\Lambda}_3^{k+1} &= \mathbf{\Lambda}_3^k + \mathbf{E}^{k+1} - \mathbf{Q}^{k+1}\end{aligned}$$

8: Update penalty parameter: $\mu_3 = \min(\mu_{max}, \rho\mu_3)$
9: Update $k : k = k + 1$.
10: **end while**
11: **return** $\hat{\mathbf{X}} = \mathbf{X}^k$, $\hat{\mathbf{E}} = \mathbf{E}^k$.

In the IALM algorithm, both Steps 4 and 6 are the ridge regression problems. Although Steps 3 and 5 are nonlinear convex optimization problems, fortunately, they both have closed-form solutions. Step 3 can be solved by the singular value thresholding operator $\mathcal{D}_\kappa(\cdot)$ [55] in the following lemma, and Step 5 can be solved by the soft-shrinkage thresholding operator $S_{\lambda_3/\mu_3}(\cdot)$ in [48].

Lemma 3: Consider the following optimization problem:

$$\mathbf{\Theta}^* = \arg\min_{\mathbf{\Theta}} \kappa_3\|\mathbf{\Theta}\|_* + \frac{1}{2}\|\mathbf{\Theta} - \mathbf{C}\|_F^2 \tag{37}$$

where the notation in (37) is the same with Lemma 2. If the rank of the matrix $\mathbf{C}$ is r, the singular value decomposition of $\mathbf{C}$ is

$$\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \mathbf{\Sigma} = \text{diag}\left(\{\pi_i\}_{1\leq i\leq r}\right) \tag{38}$$

where $\mathbf{U} \in \mathbb{R}^{m\times r}$ and $\mathbf{V} \in \mathbb{R}^{n\times r}$ are matrices with orthogonal columns and the singular values $\pi_i (1 \leq i \leq r)$ are positive. For each parameter $\kappa_3 \geq 0$, we define the singular thresholding operator $\mathcal{D}_{\kappa_3}(\cdot)$ as

$$\begin{aligned}\mathcal{D}_{\kappa_3}(\mathbf{C}) &:= \mathbf{U}\mathcal{D}_{\kappa_3}(\mathbf{\Sigma})\mathbf{V}^T \\ \mathcal{D}_{\kappa_3}(\mathbf{\Sigma}) &:= \text{diag}\left(\{(\pi_i - \kappa_3)_+\}_{1\leq i\leq r}\right)\end{aligned} \tag{39}$$

where $(\pi_i - \kappa_3)_+ = \max(\pi_i - \kappa_3, 0)$. Then, the optimal solution of $\mathbf{\Theta}^*$ for (37) is

$$\mathbf{\Theta}^* = \mathcal{D}_{\kappa_3}(\mathbf{C}).$$

Remark 3: Convergence analysis: For IALM, which is a variation of exact augmented Lagrangian method (ALM), its convergence has already been well studied when the number of blocks (i.e., unknown matrix variables) is, at most, two [57]. However, up to now, it is still difficult to generally ensure the convergence of IALM with three or more blocks. Similar to the case in [40], because there are four blocks $\mathbf{X}$, $\mathbf{E}$, $\mathbf{P}$, and $\mathbf{Q}$ in Algorithm 3 and the objective function of (34) is nonsmooth, it is difficult to prove the convergence of our proposed algorithm theoretically. In [40], the authors have introduced some conditions, which guarantees the convergence of the algorithm to some extent. Moreover, as illustrated in [57], the IALM is known to generally perform well in reality. In practice, if the dictionary size and the parameters are appropriately chosen, we observe that the proposed IALM algorithm convergences when the preset maximum iteration is reached.

VI. Experiments

In this section, first, we summarize several objective criteria for evaluating the performance of different abundance estimation algorithms for HSI. Second, the experiment environment and the parameter settings for all the involved algorithms are described. Third, we demonstrate that the proposed algorithms with ASC show better performance than without ASC. Fourth, we demonstrate the proposed SMV-ADMM method on various synthetic data sets. Fifth, we validate that our LRR model makes better estimation of the abundances for synthetic bilinear mixed images. Finally, our proposed methods are demonstrated on a real hyperspectral image.

A. Metrics for Performance Evaluation

The quality of the abundance estimation strategy for synthetic images can be measured by comparing the estimated and actual abundances using the root-mean-square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{nR}\sum_{i=1}^{n}\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2} \tag{40}$$

where $\boldsymbol{x}_i$ and $\hat{\boldsymbol{x}}_i$ are the actual and estimated abundance vectors of the ith pixel of the image and n is the number of pixels. Moreover, the signal-to-reconstruction error (SRE) introduced in [29], which provides a better measurement on the relationship between the power of error and signal, is defined as

$$\text{SRE} = 10\log_{10}\left(\frac{\text{E}\left[\|\mathbf{X}\|_2^2\right]}{\text{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2\right]}\right). \tag{41}$$

In the case of real hyperspectral images, as reported in [1] and [23], the reconstruction error (RE) and the spectral angle mapper (SAM) are often used to roughly estimate the quality of an abundance estimation algorithm. RE is defined as

$$\text{RE} = \sqrt{\frac{1}{nL}\sum_{i=1}^{n}\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2} \tag{42}$$

TABLE I
PARAMETER SETTING FOR ALL ALGORITHMS IN THE EXPERIMENTS

Algorithm	Parameters
FCLS [16]	$\lambda_f = 1 \times 10^{-3}$, $\mu_f^0 = 0.02$, $Iter = 500$, $tol = 10^{-4}$
CSUnSAL+ [46]	$\lambda_1 = 2 \times 10^{-3}$, $\mu_1^0 = 0.02$, $Iter = 500$, $tol = 10^{-4}$
GDA [25]	Not Applicable
SK-Hype [27]	$\lambda_k = 5 \times 10^{-3}$, $Iter = 15$
SMV-ADMM	$\lambda_1' = 2 \times 10^{-3}$, $\mu_1^0 = 0.02$, $\delta_1 = 0.3$, $Iter = 500$, $tol = 10^{-4}$
MMV-ADMM	$\lambda_2 = 2 \times 10^{-3}$, $\mu_2^0 = 0.02$, $\delta_2 = 0.2$, $Iter = 500$, $tol = 10^{-4}$
LRR	$\lambda_3 = 1 \times 10^{-1}$, $\mu_3^0 = 10$, $\delta_3 = 0.2$, $Iter = 500$

where L is the number of spectral bands and $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ are the measured and estimated spectra for the ith pixel. Additionally, SAM is an alternative measure to estimate the performance

$$\text{SAM} = \frac{1}{n}\sum_{i=1}^{n}\theta[\mathbf{y}_i, \hat{\mathbf{y}}_i] \tag{43}$$

where

$$\theta[\mathbf{y}_i, \hat{\mathbf{y}}_i] = \arccos\left(\frac{\langle \mathbf{y}_i, \hat{\mathbf{y}}_i\rangle}{\|\mathbf{y}_i\|_2\|\hat{\mathbf{y}}_i\|_2}\right) \tag{44}$$

and $\arccos(\cdot)$ is the inverse cosine operator.

However, as it was pointed out in [27], without ground-truth information on the abundances for the real HSI data, that the reconstruction quality measured by RE or SAM is not necessarily proportional to the quality of the abundance estimation. Therefore, the estimation performance cannot merely be judged by the results of RE and SAM.

B. Experiment Parameter Setting

All the following simulations are executed on a computer with a 3.4-GHz Intel quad-core i7 processor and 8-GB 1600-MHz RAM. We compare the proposed algorithm with four other abundance estimation algorithms: the fully constraint least squares (FCLS) [16] and the algorithm of nonnegative constraint sparse unmixing by variable splitting and augmented Lagrangian (CSUnSAL+) [46], which are dedicated to LMM, and the gradient descent algorithms (GDAs) in [23] and [25] and the kernel-based hyperspectral unmixing algorithm (named SK-Hype) [27], which are designed for nonlinear models. The parameter settings for all the algorithms are shown in Table I. In particular, we have adopted the algorithm in [46] to solve the FCLS. $Iter$ denotes the maximum iteration, and tol is the stopping tolerance. λ_f and λ_k are the regularization parameters for FCLS and SK-Hype, respectively. The kernel used for SK-Hype is a second-order polynomial kernel. μ_f^0, μ_1^0, μ_2^0, and μ_3^0 are the initializations for the Lagrangian multipliers in the corresponding algorithms. Because the algorithms are not very sensitive to the selection of the regularization parameters, these parameters are set to the same values for all the experiments.

C. Comparisons of the Proposed Algorithms With and Without ASC

In the first experiment, we show that the proposed SMV-ADMM, MMV-ADMM, and LRR algorithms show significantly better performance with ASC than without ASC. First, we randomly select 12 endmembers from a selected USGS library[2] to build up the dictionary $\mathbf{A}$ and the composite dictionary $\mathbf{M}$ according to MGBM. Then, we generate 100 mixed pixels based on MGBM. In each pixel, three endmembers in $\mathbf{A}$ are randomly selected to generate the mixture and corrupted by the Gaussian noise with SNR $= 40$ dB, where the abundance vectors are generated according to the Dirichlet distribution following the ASC and ANC. We set the bilinear coefficients as $\zeta_{ij}^l = \gamma_{ij}^l x_i^l x_j^l (1 \le i \le j \le R)$ for some γ_{ij} uniformly distributed in [0, 1]. We compare the proposed nonnegative SMV-ADMM, MMV-ADMM, and LRR on the synthetic data with and without ASC by varying the parameters δ_1, δ_2, and δ_3 from 1×10^{-3} to 10. For the MMV-ADMM and LRR, we test the algorithms on every 10 pixels and then calculate the mean values of SRE and RMSE. The comparison results are shown in Fig. 3.

From the results in Fig. 3, we can see that the abundance estimation performances are further enhanced by enforcing the ASC. Moreover, it should be noticed that the parameter $\delta_i (1 \le i \le 3)$ should be chosen appropriately. If δ_i is too small, it will weaken the strength of ASC; if it is too large, it might overwhelm the data fidelity term and cause the performance to decrease. Empirically, δ_i is chosen in the range [0.1, 1], and more theoretical analysis is left for our future work.

D. Experiment I (Single-Pixel Abundance Estimation on Synthetic Data)

In this section, in the first experiment, we compare the proposed SMV-ADMM method with state-of-the-art algorithms based a selected USGS library. In the second experiment, we show that the proposed method is robust to endmember spectral variability based on the labeled Pavia University data set.

1) Comparison With State-of-the-Art Method: The performance of the proposed SMV-ADMM algorithm is evaluated on four synthetic images generated by the LMM and three different bilinear models. Each image contains 500 pixels. First, R pure endmembers are randomly extracted from a selected USGS library $\mathbf{W} \in \mathbb{R}^{224\times 498}$ to construct the dictionary $\mathbf{A} \in \mathbb{R}^{224\times R} (R \ll 498)$. The reflectance values are measured for 224 spectral bands distributed uniformly in the interval $0.4 - 2.5$ μm. We choose $R = 12$ and 24 endmembers to generate two different dictionaries, respectively (see Fig. 4). When $R = 12$, the matrix defined in (17) is $M \in \mathbb{R}^{224\times 90}$; when $R = 24$, $M \in \mathbb{R}^{224\times 324}$. Thus, we can test the performance of our

[2] http://www.lx.it.pt/bioucas/code/sunsaldemo.zip

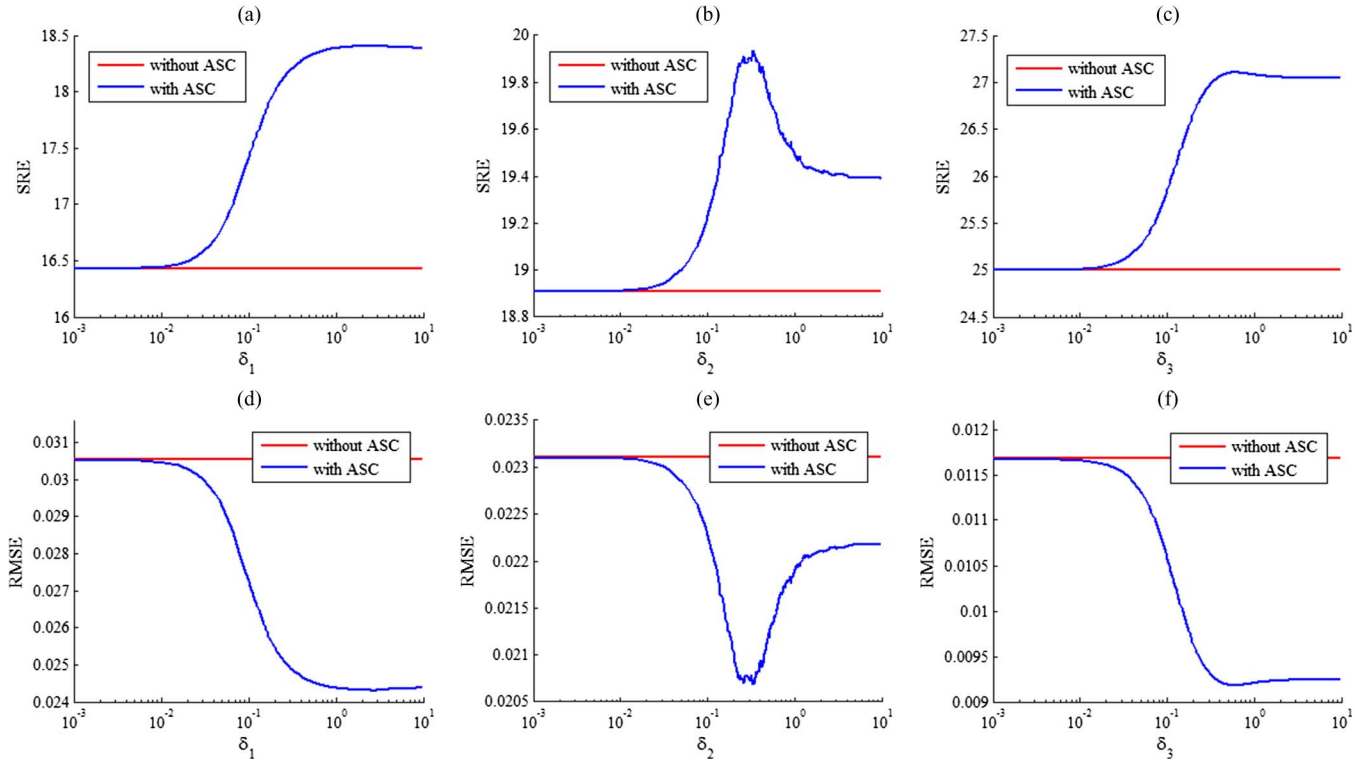


Fig. 3. Comparison of SRE and RMSE for the proposed SMV-ADMM, MMV-ADMM, and LRR algorithms with and without the ASC.

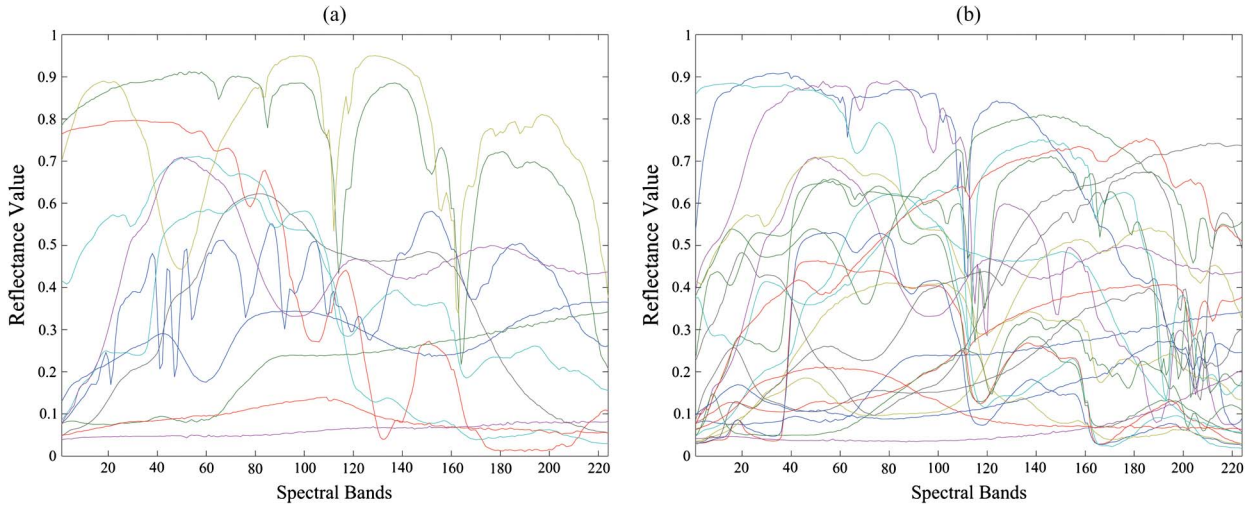


Fig. 4. Two dictionaries of sizes (a) $R = 12$ and (b) $R = 24$ from the selected USGS library.

proposed algorithm for both over- and underdetermined cases. Each pixel is a mixture of r endmembers from the dictionary $\mathbf{A}$, where r is an integer uniformly distributed in $\{1, \ldots, 6\}$. The size of the dictionary and the number of mixtures are chosen to mimic the real-life scenarios. In practice, the target hyperspectral image usually contains 10 to 15 different pure endmembers, but there might be only a few pure endmembers present in each pixel individually.

To generate the synthetic image, the abundance vectors $\mathbf{x}_l (l = 1, 2, \ldots, 2500)$ are randomly generated according to the Dirichlet distribution following the ASC and ANC. For the lth pixel, the parameters ζ_{ij}^l in MGBM are set as $\zeta_{ij}^l = \gamma_{ij}^l x_i^l x_j^l (1 \le i \le j \le R)$, where x_i^l and x_j^l are the abundances for the ith and jth endmembers and the variable γ_{ij}^l is randomly generated according to the uniform distribution over [0.5, 1]. The parameter b in PPNMM is uniformly distributed between [0, 0.5]. We corrupt each pixel by a white Gaussian noise with SNR $= 40$ dB (SNR $:= \|\mathbf{A}\boldsymbol{x}\|_2^2 / \|\mathbf{n}\|_2^2$). As in the real world, hyperspectral images are corrupted by band-correlated noise, and we also demonstrate the proposed method on pixels corrupted by colored noise with SNR $= 40$ dB for dictionary size $R = 12$. We use a first-order autoregressive filter $v(t) = 0.9v(t) + n(t-1)$ to generate a bandwise correlated noise, where $n(t)$ is a sequence of independent white Gaussian noise.

TABLE II
COMPARISON OF FIVE ABUNDANCE ESTIMATION ALGORITHMS ON LMM AND THREE DIFFERENT BILINEAR MODELS WITH DICTIONARY SIZE $R = 12$ AND WHITE GAUSSIAN NOISE

Criteria	Model	Linear Algorithm		Nonlinear Algorithm		
		FCLS [16]	CSUnSAL+ [46]	GDA [23]	SK-Hype [27]	SMV-ADMM
SRE (dB)	LMM	**36.2207**	33.0667	36.1686	27.9674	33.4288
	FM	9.1978	14.2341	11.2274	19.0686	**24.0441**
	PPNMM	8.4389	12.1792	14.8256	17.7874	**22.5689**
	MGBM	4.4667	6.8863	6.1276	14.3149	**20.1900**
RMSE ($\times 10^{-2}$)	LMM	**0.3298**	0.4742	0.3318	0.8529	0.4548
	FM	7.3903	4.1385	5.8503	2.3720	**1.3377**
	PPNMM	8.0766	5.2506	3.8716	2.7529	**1.5876**
	MGBM	12.6640	9.5849	10.4598	4.0753	**2.0721**
RE ($\times 10^{-2}$)	LMM	**0.0883**	0.0995	0.0945	0.1332	0.1069
	FM	3.7846	0.8129	2.4373	0.3374	**0.1635**
	PPNMM	7.3660	1.1486	0.8583	0.4052	**0.1711**
	MGBM	15.8858	2.4218	14.2949	0.8057	**0.1812**
SAM ($\times 10^{-2}$)	LMM	0.1378	0.1588	**0.1307**	0.1878	0.1826
	FM	3.3896	1.1363	2.3878	0.4832	**0.2173**
	PPNMM	4.2059	1.3075	0.7991	0.5133	**0.2241**
	MGBM	8.3694	2.5086	6.9952	0.8894	**0.2083**

TABLE III
COMPARISON OF FIVE ABUNDANCE ESTIMATION ALGORITHMS ON LMM AND THREE DIFFERENT BILINEAR MODELS WITH DICTIONARY SIZE $R = 12$ AND COLORED NOISE

Criteria	Model	Linear Algorithm		Nonlinear Algorithm		
		FCLS [16]	CSUnSAL+ [46]	GDA [23]	SK-Hype [27]	SMV-ADMM
SRE (dB)	LMM	**33.7152**	31.0981	33.6897	26.9026	31.4693
	FM	9.1736	14.1991	11.1618	18.8529	**23.7791**
	PPNMM	8.4377	12.2291	14.5667	17.7527	**22.0611**
	MGBM	4.4655	6.9151	6.1609	14.3233	**20.0478**
RMSE ($\times 10^{-2}$)	LMM	**0.4401**	0.5947	0.4413	0.9641	0.5699
	FM	7.4109	4.1553	5.8947	2.4317	**1.3791**
	PPNMM	8.0778	5.2206	3.9888	2.7640	**1.6831**
	MGBM	12.6658	9.5533	10.4200	4.0714	**2.1063**
RE ($\times 10^{-2}$)	LMM	**0.1229**	0.1394	0.1254	0.1768	0.1473
	FM	3.7911	0.8166	2.4478	0.3513	**0.2002**
	PPNMM	7.3663	1.1513	0.7817	0.4179	**0.2046**
	MGBM	15.8858	2.4230	14.3143	0.8115	**0.2120**
SAM ($\times 10^{-2}$)	LMM	**0.2111**	0.2417	0.2021	0.2813	0.2769
	FM	3.4010	1.1565	2.4198	0.5314	**0.2974**
	PPNMM	4.2141	1.3263	0.7820	0.5589	**0.3002**
	MGBM	8.3750	2.5151	6.9952	0.9096	**0.2700**

The proposed algorithm is compared with FCLS, CSUnSAL +, GDA, and SK-Hype. It should be noticed that the GDA in [23] and [25] depends on the exact model assumption. As such, we need to manually modify the algorithms to fit different model assumptions for FM, PPNMM, and MGBM.

From the result in Tables II–IV, we can see the following.

1) For various BMMs, the proposed nonnegative SMV-ADMM algorithm shows significantly improved performance compared with state-of-the-art algorithms no matter whether the system is overdetermined $M \in \mathbb{R}^{224\times 90}$ or underdetermined $M \in \mathbb{R}^{224\times 324}$.
2) Comparing the results in Tables II and III, the proposed SMV-ADMM shows similar performance for pixels corrupted by white Gaussian noise and colored noise.
3) For the synthetic data generated by LMM, our method shows worse results compared with FCLS based on all the criteria, but for the data generated by BMMs, our method outperforms FCLS. Therefore, we can see that the introduced bilinear dictionary of our method is neither "overfitting" the data nor modeling the sensor noise.

We conclude that the introduced bilinear dictionary in our proposed SMV-ADMM algorithm is very effective for detecting and getting rid of the bilinear components in BMMs.

2) Experiment II (Endmember Spectral Variability Versus Nonlinear Mixture): In this experiment, we show that the proposed bilinear abundance estimation method is robust to endmember spectral variability [56] and is not intended for "overfitting" or "memorizing" the data. To demonstrate this, we use the well-labeled Pavia University data set,[3] which contains nine manually labeled materials, including shadow. Each material has many labeled pixels in the data set, and the pixels that are labeled as the same material have small spectral variabilities at different spatial locations due to the varying condition of scene components and differential illumination [56].

First, we choose six different materials (see Fig. 5) to generate the linear and bilinear dictionaries. For each material, a pure endmember is estimated by the mean value of all the

[3]http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

TABLE IV
COMPARISON OF FIVE ABUNDANCE ESTIMATION ALGORITHMS ON LMM AND THREE DIFFERENT BILINEAR MODELS WITH DICTIONARY SIZE $R = 24$ AND WHITE GAUSSIAN NOISE

Criteria	Model	Linear Algorithm		Nonlinear Algorithm		
		FCLS [16]	CSUnSAL+[46]	GDA [23]	SK-Hype [27]	SMV-ADMM
SRE (dB)	LMM	27.8323	23.9878	**27.9793**	17.9945	27.2918
	FM	11.1918	15.9883	11.7764	15.8477	**20.7901**
	PPNMM	10.2880	13.8383	11.8308	14.1336	**19.9971**
	MGBM	6.3574	8.8930	7.8705	10.9300	**17.6082**
RMSE ($\times 10^{-2}$)	LMM	0.5931	0.9233	**0.5832**	1.8409	0.6312
	FM	4.0261	2.3178	3.7641	2.3556	**1.3334**
	PPNMM	4.4998	2.9900	3.7674	2.8901	**1.4714**
	MGBM	7.0182	5.2414	5.8962	4.1457	**1.9217**
RE ($\times 10^{-2}$)	LMM	0.1149	0.1245	**0.0949**	0.0240	0.1321
	FM	1.4598	0.4278	1.9303	0.2565	**0.2010**
	PPNMM	2.6250	0.7163	0.6806	0.2904	**0.2577**
	MGBM	5.6073	1.4365	8.6625	0.86625	**0.2701**
SAM ($\times 10^{-2}$)	LMM	0.3015	0.3275	**0.2651**	0.4289	0.4064
	FM	2.0012	0.8351	2.2356	0.5189	**0.4654**
	PPNMM	2.8825	1.0727	1.1489	0.5765	**0.5092**
	MGBM	5.3896	1.9078	4.6970	0.7984	**0.4637**

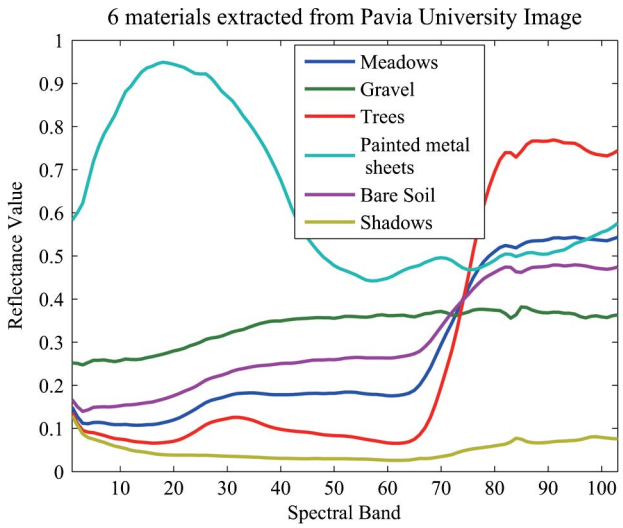


Fig. 5. Mean spectral value for the six different labeled materials in Pavia University hyperspectral image. The endmembers with bold solid line are chosen for generating the mixture in the experiment.

pixels labeled as the same material, and the bilinear endmember is generated by the Hadamard product of two mean pure endmembers. Second, 20 mixed pixels are generated based on both the LMM and MGBM. For both cases, the abundances are set the same, randomly generated using Dirichlet distribution. The bilinear abundance for MGBM is generated in the same way as in Experiment I. To account for the spectral variability, every synthetic pixel is a mixture of two pure pixels where each one is a randomly selected pixel from the pool of pixels that are labeled as the same material. The synthetic pixels are also corrupted by a Gaussian white noise of SNR $= 40$ dB. We use the proposed SMV-ADMM to estimate the abundances for the mixed pixels with spectral variability for both the LMM and MGBM.

The estimated linear and bilinear abundances along with the ground truth are shown in Fig. 6. From the results, we can see that the estimated bilinear abundances by the proposed SMV-ADMM method for the LMM case are nearly zero, while for the MGBM case, our method can still approximately predict the positions and values for the bilinear abundances even with the occurrence of endmember spectral variability for each mixed pixel. Therefore, we conclude that the proposed approach is robust to spectral variability and is able to accurately distinguish between bilinearly mixed data and pixels that are linear mixtures of endmembers with shadow and spectral variabilities.

E. Abundance Estimation by Jointly Sparse and LRR Method on Synthetic Data

In this section, we use two experiments, demonstrating the following: 1) The proposed bilinear dictionary can effectively deal with the bilinear term in the BMM; 2) the joint-sparse regression fails when pixels consist of different materials; 3) by using the spatial information, the joint sparsity and LRR model can improve the performance of the abundance estimation; and 4) the LRR model can better capture the HSI data structure than the joint-sparsity model.

1) Experiment I (Effectiveness of the Bilinear Dictionary): In the first experiment of this section, we use a toy example to demonstrate the effectiveness of the bilinear dictionary and the deficiency of the joint-sparsity model. We randomly select 12 pure endmembers from the USGS library [see Fig. 7(a)] and use four of them to generate our mixed pixels. Under the assumption that the materials are homogeneous (all pixels have the same endmembers), 10 pixels are generated by MGBM in the same way as in the previous section. The first 5 pixels are mixed by the 1–4th endmembers, and the rest are mixed by the 5–8th endmembers. The data are corrupted by the white Gaussian noise with SNR $= 40$ dB. The estimated linear and bilinear abundances are depicted in Fig. 2, where each column denotes one pixel in each subimage and each row indicates one pure endmember in the dictionary. From the results in Fig. 2 and Table V, we can see the following.

1) The proposed methods can well approximate the bilinear term generated by MGBM.

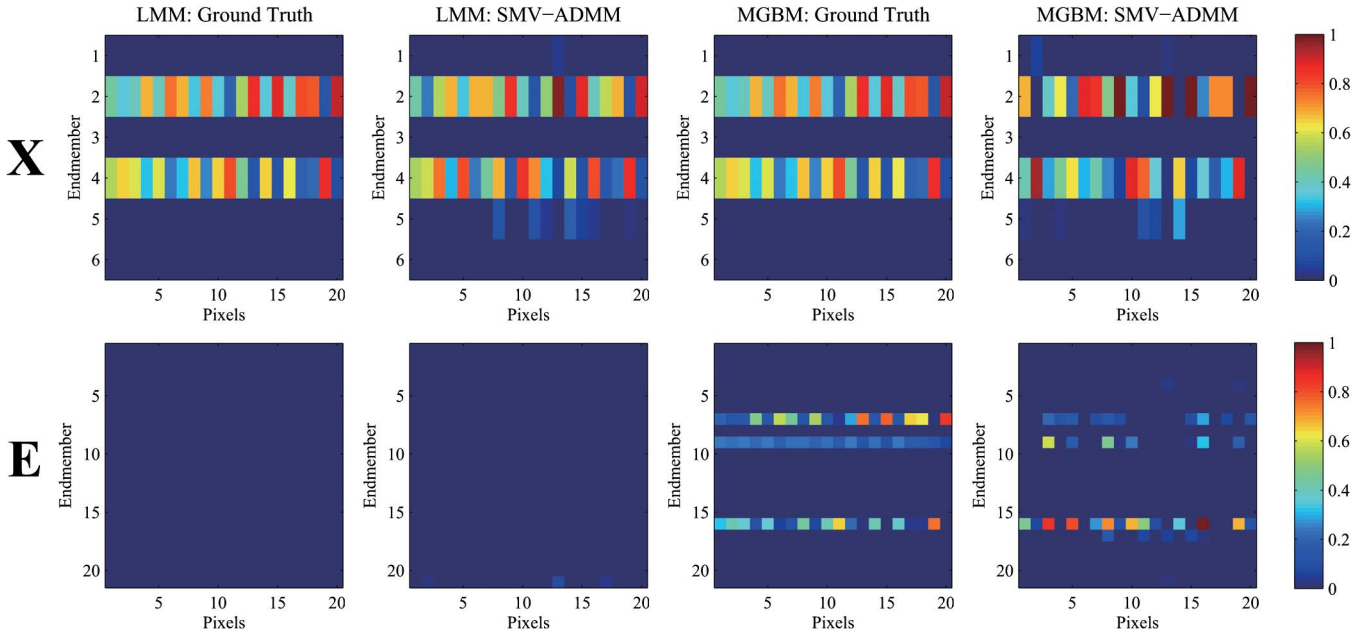


Fig. 6. Comparisons of the ground truth with the estimated abundance matrix $\mathbf{X}$ and the bilinear coefficient matrix $\mathbf{E}$ by SMV-ADMM in Experiment II. The first row shows the abundance matrices, and the second row shows the bilinear coefficient matrices. From the left column to the right are the ground truth for LMM, estimation for LMM, ground truth for MGBM, and estimation for MGBM.

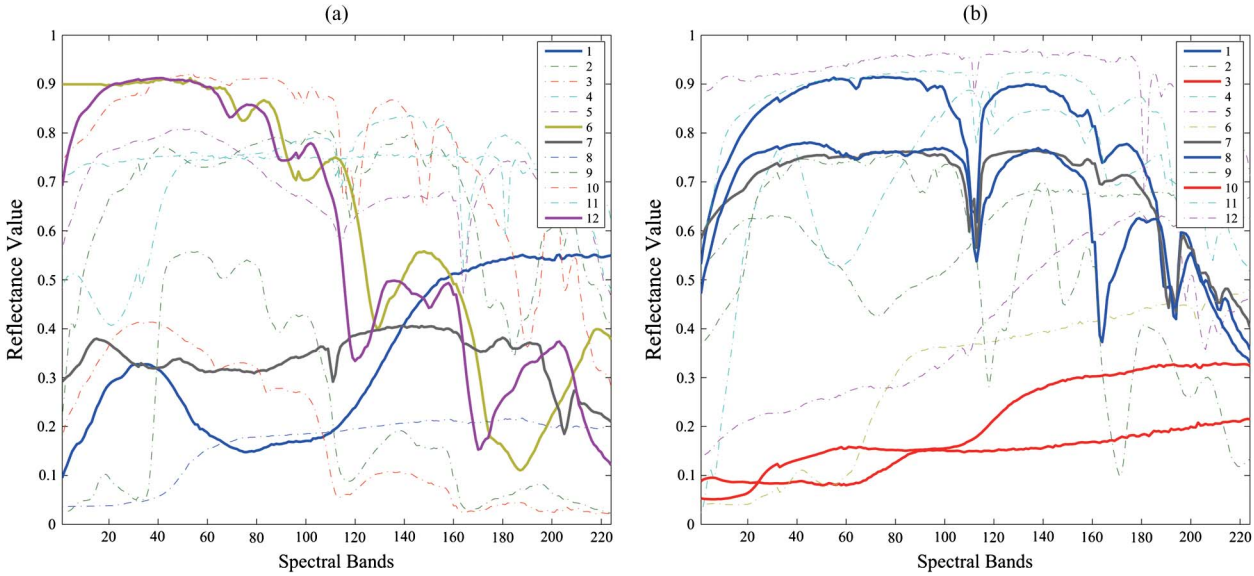


Fig. 7. Two endmember libraries of size $R = 12$ for Experiments I and II in Section VI-D, respectively. The endmembers with bold solid line are chosen for generating the mixture in the experiment.

TABLE V
COMPARISON OF UNMIXING RESULTS FOR FIG. 2

Algorithm	SMV -ADMM	MMV -ADMM	Low Rank Representation
SRE (dB)	13.1909	10.1044	**18.1065**
RMSE ($\times 10^{-2}$)	3.8213	5.4518	**1.9317**
RE ($\times 10^{-3}$)	4.4832	7.1348	**4.2123**
SAM ($\times 10^{-3}$)	7.9967	12.5619	**7.7046**

2) The joint-sparse regression fails when pixels within a neighborhood contain different sets of endmembers, and the LRR method has the best performance since it does not enforce a strict row-sparsity constraint.

2) Experiment II (Effectiveness of the LRR): In this experiment, we demonstrate the effectiveness of the joint-sparsity model and LRR method on a synthetic hyperspectral image. Again, we randomly select 12 pure endmembers from the USGS library to construct the endmember dictionary [see Fig. 7(b)] and the corresponding second-order bilinear dictionary. Five endmembers are then randomly selected from the endmember dictionary to generate a 150 × 150 image by MGBM as illustrated in Fig. 8(a). Each background pixel is generated by a bilinear mixture of five endmembers with the same abundance and bilinear coefficients. We generate a total number of 25 blocks of size 20 × 20 arrayed in five rows,

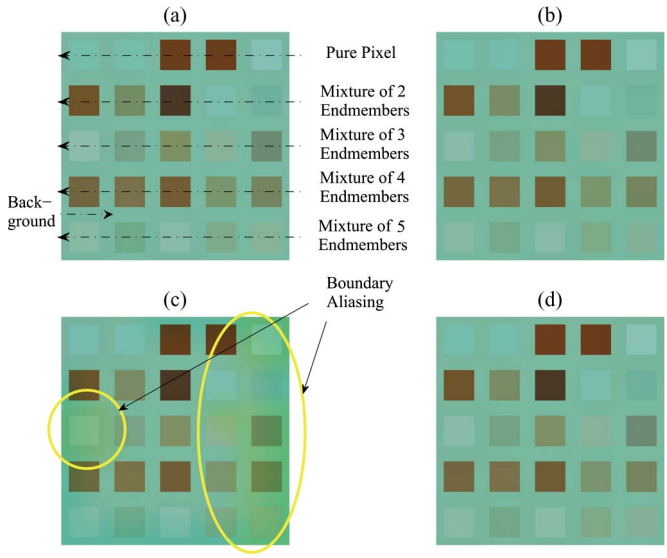


Fig. 8. Original and the reconstructed synthetic false-color hyperspectral images of the 1th, 112th, and 224th bands. (a) Synthetic image of size 150 × 150 generated according to MGBM by a small dictionary of 12 pure endmembers extracted from the USGS library. In this image, the background is generated by a mixture of five pure endmembers randomly selected from the dictionary, each 20 × 20 block is randomly mixed by 1–5 endmembers in the background, and every pixel is corrupted by Gaussian white noise with SNR = 40 dB. (b) Reconstructed image from SMV-ADMM algorithm. (c) Reconstructed image from MMV-ADMM algorithm. (d) Reconstructed image from our LRR method.

TABLE VI
COMPARISON OF UNMIXING RESULTS FOR FIG. 8

Algorithm	SMV -ADMM	MMV -ADMM	Low Rank Recovery
SRE (dB)	12.9585	14.1834	**20.2845**
RMSE (10^{-2})	1.0680	0.9276	**0.4595**
RE (10^{-3})	5.4922	5.4531	**5.2966**
SAM (10^{-2})	1.0492	1.0268	**1.0051**

where the blocks in each row have the same number (varying from one to five) of endmembers randomly selected from the five endmembers but mixed with different abundances. In each row, there are five blocks positioned in parallel, and within each block, the pixels are generated according to MGBM with the same abundances. Throughout the entire HSI image, each pixel is corrupted by the white Gaussian noise with SNR = 40 dB.

As in our previous experiments, we have already demonstrated the effectiveness of the proposed SMV-ADMM algorithm for the BMMs. Therefore, in this experiment, we only compare our proposed SMV-ADMM, MMV-ADMM, and LRR algorithms. The reconstructed images and the estimated abundances for the proposed algorithms are shown in Figs. 8 and 9, respectively. The results of the abundance estimation error are demonstrated in Table VI. From the results, we conclude the following.

1) By taking advantage of the spatial information, the joint-sparsity model can reduce the estimation error to some extent. However, it causes some *aliasing artifacts* for the pixels on the boundaries between different materials.
2) The LRR method further reduces the estimation error significantly. Unlike the joint-sparsity model, it does not cause severe *aliasing artifacts*.

F. Real Data

In this section, we evaluate the performance of the proposed method when executed on a real hyperspectral image. The data under consideration are the well-known Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite image (refer to [58] for a detailed description) over the Cuprite mining region in NV, USA, which is available online in reflectance units.[4] The portion that we use in this experiment is the image of the sector labeled as "f970619t01p02_r02_sc02.a.rfl" of size 512 × 614 in the data set [see Fig. 10(a)]. The scene comprises 224 spectral bands between 0.4 and 2.5 μm, with a nominal spectral resolution of 10 nm. Prior to the analysis, bands 1–2, 105–115, 150–170, and 223–224 are removed due to water absorption and low SNR in those bands, leaving a total of 188 spectral bands available [29].

Before performing abundance estimation, our first step is to determine the pure endmembers. This can be accomplished by the VCA algorithm. By assuming the existence of pure pixels in the Cuprite image and based on the geometry of convex sets, the VCA algorithm exploits the fact that endmembers occupy the vertices of a simplex. It iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined and selects a new endmember signature that corresponds to the extreme of the projections. The VCA can efficiently extract endmembers with high accuracy.

However, it should be noticed that the VCA algorithm relies on the assumption that the pixels in the data are linearly mixed. However, in our case, we are dealing with nonlinear mixtures, and as explained in [23] and [25], we can still apply it to extract endmembers for bilinear models where only small nonlinearities occur in the data. On the other hand, because VCA is an unsupervised algorithm, we need to decide the number of pure endmembers to extract, which is not a trivial task. A too small number will not yield good estimation results since the retrieved endmembers themselves will be mixtures of several actual pure materials. A too large number of endmembers will result in many unidentifiable endmembers, mostly corresponding to spectral noise, and multiple endmembers corresponding to the same pure material. In our experiment, we empirically extract 12 pure endmembers by the VCA algorithm, where Fig. 10(b) shows the extracted pure endmembers for one experiment.

After the pure endmembers are extracted, we construct the pure endmember dictionary and the associated bilinear dictionary. Then,the abundance estimation is conducted on a subimage of size 200 × 200 from the original 512 × 614 data [the region within the yellow rectangle in Fig. 10(a)]. We compare our proposed SMV-ADMM, MMV-ADMM, and LRR algorithms with FCLS, CSUnSAL+, GDA, and SK-Hype with different model assumptions. Additionally, we perform our SMV-ADMM, MMV-ADMM, and LRR on the proposed MGBM compared with GBM, which is accomplished by a smaller bilinear dictionary excluding the self-reflection components $a_i \odot a_i (1 \leq i \leq R)$ in (16).

Because VCA is initialization dependent, we run the experiment 15 times to show the stability of the proposed algorithms.

[4] http://aviris.jpl.nasa.gov/html/aviris.freedata.html

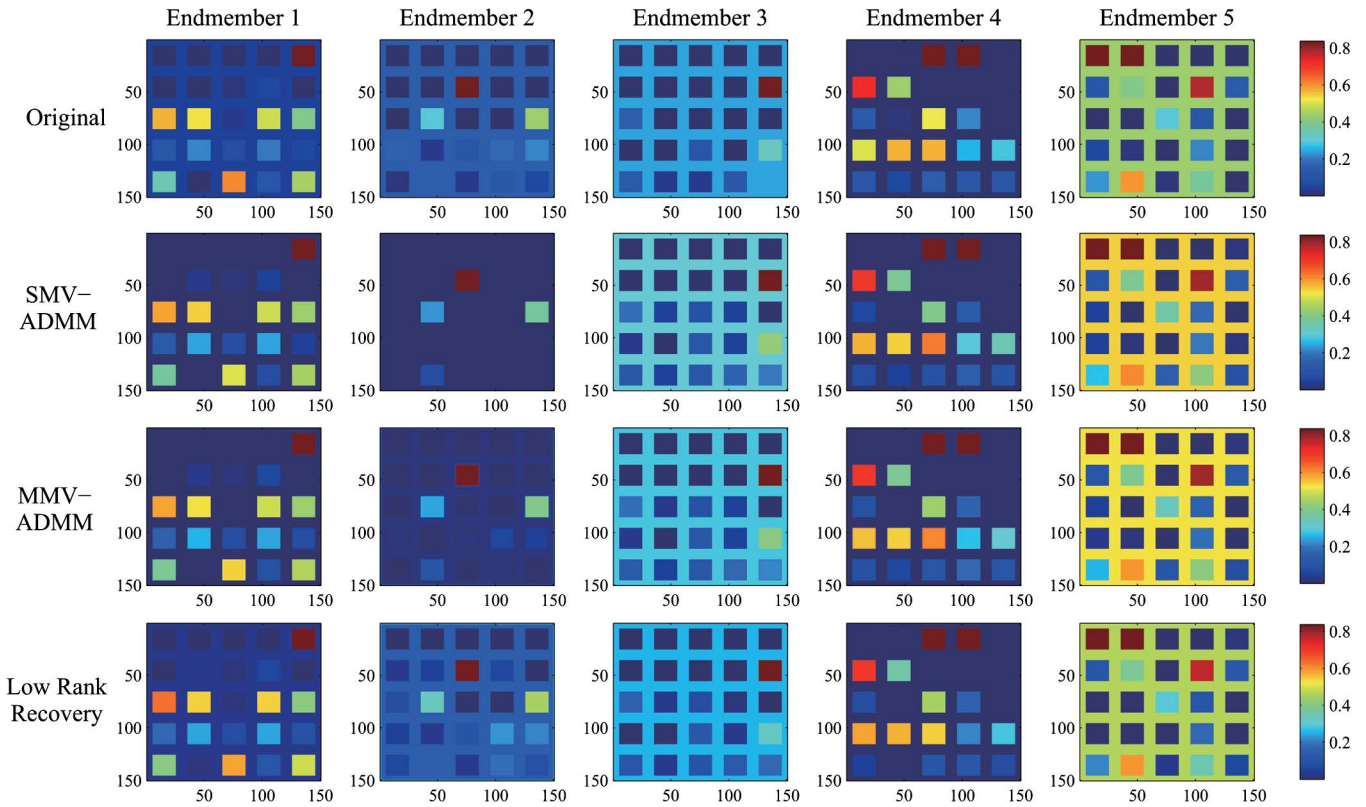


Fig. 9. Comparison of abundance estimation result on a synthetic image. The first row shows the abundance of five different endmembers in the original image, and the remaining rows show the estimated abundance results for the proposed SMV-ADMM, MMV-ADMM, and LRR algorithms, respectively. Each column shows the original and the estimated image of abundances for one mixed pure endmember.

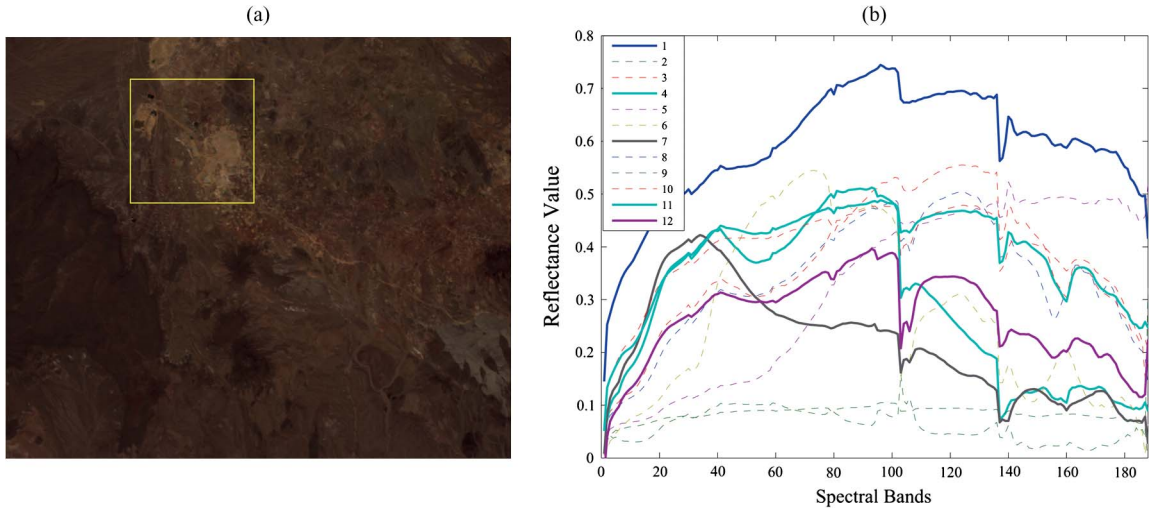


Fig. 10. Real data used in the final experiment. (a) False-color image of the AVIRIS Cuprite data set in the file "f970619t01p02_r02_sc02.a.rfl." The bands used as RGB channels are bands (12, 22, 42) of the original 224-band image. (b) Twelve pure endmembers extracted by VCA for one experiment are displayed. The endmembers with bold line are the selected ones, with their abundances shown in Fig. 11.

The results of RE, SAM, and their associated means and variances are shown in Table VII, where μ_{RE} and σ_{RE} are the mean and variance for RE and μ_{SAM} and σ_{SAM} are the mean and variance for SAM. Obviously, the proposed algorithms based on MGBM show much lower reconstructed errors compared with other approaches. However, it should be noticed that the RE and SAM here are only reported as complementary information as explained in Section VI-A.

The estimated abundances for one of the 15 experiments are reported in Fig. 11, where the proposed SMV-ADMM, MMV-ADMM, and LRR are based on MGBM. We can see that all the algorithms generate abundance maps with similar patterns, but our proposed algorithms produce seemingly more meaningful abundances. Specifically, for Endmember 1, the abundance maps generated by our algorithms are sharper and clearer; for Endmember 12, the proposed methods could find

TABLE VII
COMPARISON OF THE RECONSTRUCTION ERROR FOR ALL ALGORITHMS ON THE REAL HYPERSPECTRAL IMAGE WITH ABUNDANCE SHOWN IN FIG. 11

Model	Algorithm	μ_{RE} ($\times 10^{-3}$)	σ_{RE} ($\times 10^{-7}$)	μ_{SAM} ($\times 10^{-2}$)	σ_{SAM} ($\times 10^{-6}$)
LMM	FCLS [16]	7.5587	12.4031	2.2560	16.4715
	CSUnSAL+ [46]	6.4522	7.7908	1.9820	9.8041
PPNMM	GDA [25]	6.3501	7.8927	1.9085	9.2313
Nonlinear	SK-Hype [27]	6.0827	1.4760	1.8863	2.0917
GBM	SMV-ADMM	5.7820	11.3906	1.7974	12.6686
	MMV-ADMM	5.6314	11.9591	1.7559	13.1640
	LRR	4.1707	1.3997	1.3023	1.7020
MGBM	SMV-ADMM	5.7090	11.3740	1.7762	12.7006
	MMV-ADMM	5.5422	11.8672	1.7306	13.1384
	LRR	**3.9867**	1.1569	**1.2444**	1.3967

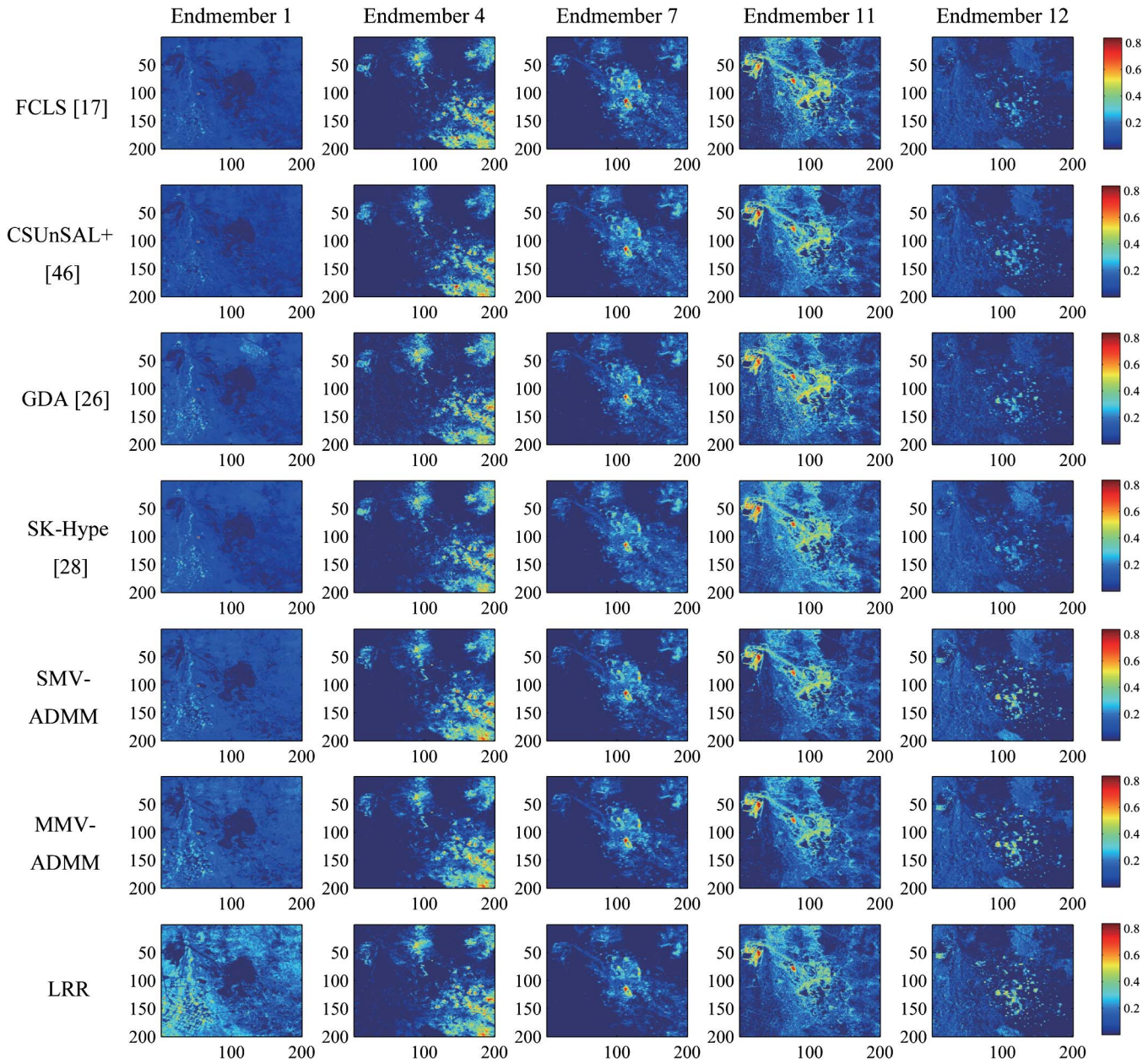


Fig. 11. Real abundance estimated by seven different algorithms for five different extracted endmembers.

the materials in the up-left corner that the other algorithms cannot. Without ground truth for the real data, more convincing validations are needed and remain for the future work. In any cases, various experiments on synthetic data in the previous sections have already demonstrated the effectiveness of our methods.

VII. Conclusion and Future Work

In this paper, we have proposed an SMV-ADMM algorithm for hyperspectral abundance estimation problem based on the BMMs, which transforms the bilinear problem into a linear problem. The MMV-ADMM and LRR algorithms are further proposed to exploit the spatial structure of the abundance vectors. Extensive simulation results demonstrate that our proposed methods significantly outperform state-of-the-art algorithms. However, the proposed methods still have limitations and drawbacks.

First, the proposed methods are dedicated to BMMs and cannot be generalized if the underling mixture mechanism is complex and does not satisfy the BMM models. Furthermore, as the size of the linear dictionary is increased, the size of the bilinear dictionary will increase dramatically, which will make the proposed algorithms impractical for large dictionaries like the case in [29]. Second, as shown in Section VI-C, the proposed methods are sensitive to the value of the ASC regularization parameter δ. Some theoretical analysis for choosing this parameter needs to be addressed. On the other hand, solving the LRR is currently computationally expensive. Therefore, a more efficient optimization algorithm for this problem is required (i.e., the linearized augmented Lagrangian method [59]).

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments on improving the quality of this paper. Moreover, the authors would also like to thank J. Chen for generously providing their code of the SK-Hype algorithm for comparison purposes.

References

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

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "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.

[3] A. Plaza, G. Martin, J. Plaza, M. Zortea, and S. Sanchez, "Recent developments in endmember extraction and spectral unmixing," in *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, S. Prasad, L. Bruce, and J. Chanussot, Eds. New York, NY, USA: Springer-Verlag, 2011, pp. 235–267.

[4] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Summaries 5th Annu. JPL Airborne Geosci. Workshop*, R. Green, Ed., Pasadena, CA, USA, 1995, vol. 1, pp. 23–26.

[5] H. Ren and C. I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.

[6] J. 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.

[7] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

[8] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[9] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, 2008, pp. 250–253.

[10] M. E. Winter, "N-finder: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE*, Denver, CO, USA, 1999, vol. 3753, pp. 266–275.

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

[12] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[13] 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.

[14] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 219–228, Jan. 2012.

[15] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained $l_p - l_2$ optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142–1146, Sep. 2013.

[16] D. C. Heinz and C. 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.

[17] P. Honeine and C. Richard, "Geometric unmixing of large hyperspectral images: A barycentric coordinate approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2185–2195, Jun. 2012.

[18] J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.

[19] N. Dobigeon, J. Y. Tourneret, and C. I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.

[20] B. Hapke, "Bidirectional reflectance spectroscopy: 1. Theory," *J. Geophys. Res.*, vol. 86, no. B4, pp. 3039–3054, Apr. 1981.

[21] Y. Altmann, N. Dobigeon, and J. Y. Tourneret, "Bilinear models for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS WHISPERS*, Lisbon, Portugal, Jun. 2011, pp. 1–4.

[22] P. Gader, D. Dranishnikov, A. Zare, and J. Chanussot, "A sparsity promoting bilinear unmixing model," in *Proc. IEEE GRSS WHISPERS*, Shanghai, China, Jun. 2012.

[23] A. Halimi, Y. Altmann, N. Dobigeon, and J. Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[24] C. Jutten and J. Karhunen, "Advances in non-linear blind source separation," in *Proc. 4th Workshop Ind. Compon. Anal. Blind Signal Separation*, Nara, Japan, Apr. 2003, pp. 245–256.

[25] Y. Altmann, A. Halimi, N. Dobigeon, and J. Y. Tourneret, "Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, Jun. 2012.

[26] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE IGARSS*, Barcelona, Spain, Jul. 2007, pp. 4041–4044.

[27] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.

[28] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Proc. SPIE Conf. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imag. XV*, Orlando, FL, USA, 2009, p. 73341M.

[29] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

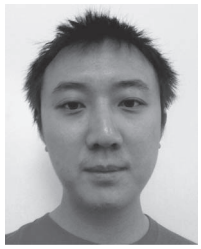
[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Nov. 2010.

[31] J. Yang and Y. Zhang, "Alternating Direction Algorithms for l_1-Problems in Compressive Sensing," Rice Univ., Houston, TX, USA, Tech. Rep. CAAM TR09-37, 2010.

[32] 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.

[33] 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.

[34] O. Eches, N. Dobigeon, and J. Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, May 2011.
[35] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[36] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
[37] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process. Special Issue Sparse Approx. Signal Image Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
[38] E. Berg and M. Friedlander, "Joint-sparse recovery from multiple measurements," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2516–2527, 2010.
[39] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low rank representation," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 663–670.
[40] 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.
[41] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York, NY, USA: Wiley, 2000.
[42] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Remote Sens. Environ.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.
[43] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer-Verlag, 2010.
[44] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc., Series B*, vol. 58, no. 1, pp. 267–288, 1996.
[45] T. Goldstein and S. Osher, "The split Bregman method for l_1 regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
[46] J. M. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS WHISPERS*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.
[47] X. Meng and M. W. Mahoney, "Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression," in *Proc. 45th STOC*, 2013, pp. 91–100.
[48] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
[49] H. Lu, X. Long, and J. Lv, "A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods," in *Proc. 14th Int. Conf. Artificial Intelligence and Statistics*, Ft. Lauderdale, FL, USA, May 2011, pp. 461–469.
[50] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, Feb. 1990.
[51] L. Zhang, H. Gao, Z. Lin, Y. Ma, X. Zhang, and N. Yu, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proc. IEEE Conf. CVPR*, Providence, RI, USA, Jun. 2012, pp. 2328–2335.
[52] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, May 2011.
[53] M. Fazel, "Matrix Rank Minimization With Applications," Ph.D. dissertation, Dept. Electr. Eng., Stanford Univ., Stanford, CA, USA, 2002.
[54] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," Univ. Illinois at Urbana-Champaign, Champaign, IL, USA, UIUC Tech. Rep., UILU-ENG-09-2215, 2009.
[55] J. Cai, J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[56] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603–1616, Apr. 2011.
[57] Y. Zhang, "Recent advances in alternating direction methods: Practice and theory," presented at the IPAM Workshop, Los Angeles, CA, USA, Oct. 11, 2010. Tutorial. [Online]. Available: https://www.ipam.ucla.edu/publications/opws2/opws2_9066.pdf
[58] F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
[59] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," in *Proc. NIPS*, Granada, Spain, Dec. 2011, pp. 61–20.



Qing Qu (S'12) received the B.Eng. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011 and the M.Sc. degree in electrical engineering from the Johns Hopkins University, Baltimore, MD, USA, in 2013.

He worked as an Intern at the U.S. Army Research Laboratory, Adelphi, MD, USA, from June 2012 to August 2013. He is currently with the Electrical Engineering Department, Columbia University. His research interests include the theory and applications of compressed sensing and sparse representations, machine learning, and large-scale optimization problems.

Nasser M. Nasrabadi (S'80–M'84–SM'92–F'01) received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from the Imperial College of Science and Technology (University of London), London, England, in 1980 and 1984, respectively.

From October 1984 to December 1984, he worked for IBM U.K. as a Senior Programmer. During 1985 to 1986, he worked with the Philips research laboratory in NY, USA, as a member of the technical staff. From 1986 to 1991, he was an Assistant Professor in the Department of Electrical Engineering at Worcester Polytechnic Institute, Worcester, MA, USA. From 1991 to 1996, he was an Associate Professor with the Department of Electrical and Computer Engineering at the State University of New York at Buffalo, Buffalo, NY, USA. Since September 1996, he has been a Senior Research Scientist (ST) with the U.S. Army Research Laboratory (ARL), Adelphi, MD, USA, working on image processing and automatic target recognition. His current research interests are in hyperspectral imaging, automatic target recognition, statistical machine learning theory, robotics, and neural network applications to image processing.

Dr. Nasrabadi has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS, SYSTEMS AND VIDEO TECHNOLOGY, and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is also a Fellow of ARL and the International Society for Optics and Photonics.

Trac D. Tran (S'94–M'98–SM'08) received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1993 and 1994, respectively, and the Ph.D. degree in electrical engineering from the University of Wisconsin, Madison, WI, USA, in 1998.

In July of 1998, he joined the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA, where he currently holds the rank of Professor. In the summer of 2002, he was an American Society for Engineering Education/ONR Summer Faculty Research Fellow at the Naval Air Warfare Center Weapons Division at China Lake, CA, USA. He is currently a regular consultant for the U.S. Army Research Laboratory in Adelphi, MD, USA. His research interests are in the field of digital signal processing, particularly in sparse representation, sparse recovery, sampling, multirate systems, filter banks, transforms, wavelets, and their applications in signal analysis, compression, processing, and communications. His pioneering research on integer-coefficient transforms and pre-/postfiltering operators has been adopted as critical components of Microsoft Windows Media Video 9 and JPEG XR—the latest international still-image compression standard International Organization for Standardization/International Electrotechnical Commission 29199-2.

Dr. Tran was the Codirector (with Prof. J. L. Prince) of the 33rd Annual Conference on Information Sciences and Systems, Baltimore, MD, USA, in March 1999. He has served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING as well as IEEE TRANSACTIONS ON IMAGE PROCESSING. He was a former member of the IEEE Technical Committee on Signal Processing Theory and Methods and is a current member of the IEEE Image Video and Multidimensional Signal Processing Technical Committee. He was the recipient of the NSF CAREER Award in 2001, the William H. Huggins Excellence in Teaching Award from Johns Hopkins University in 2007, and the Capers and Marion McDonald Award for Excellence in Mentoring and Advising in 2009.