IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 30, 2018, accepted December 9, 2018, date of publication January 3, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890278

An Outlier-Insensitive Unmixing Algorithm With Spatially Varying Hyperspectral Signatures

YAO-RONG SYU[1], CHIA-HSIANG LIN[2], (Member, IEEE), AND CHONG-YUNG CHI[1], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan
[2]Department of Electrical Engineering (Institute of Computer and Communication Engineering), National Cheng Kung University, Tainan 70101, Taiwan

Corresponding author: Chong-Yung Chi (cychi@ee.nthu.edu.tw)

This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under Grant MOST104-2221-E-007-069-MY3, and in part by the Young Scholar Fellowship Program (Einstein Program), MOST, Taiwan, under Grant MOST107-2636-E-008-003.

ABSTRACT Effective hyperspectral unmixing (HU) is essential to the estimation of the underlying materials' signatures (endmember signatures) and their spatial distributions (abundance maps) from a given image (data) of a hyperspectral scene. Recently, investigating HU under the non-negligible endmember variability (EV) and outlier effects (OE) has drawn extensive attention. Some state-of-the-art works either consider EV or consider OE, but none of them considers both EV and OE simultaneously. In this paper, we propose a novel HU algorithm, referred to as the variability/outlier-insensitive multi-convex unmixing (VOIMU) algorithm, which is robust against both EV and OE. Considering two suitable regularizers, a nonconvex minimization problem is formulated for which the perturbed linear mixing model proposed by Thouvenin *et al.*, is used for modeling EV, while OE is implicitly handled by applying a p quasi-norm to the data fitting with $0 < p < 1$. Then, we reformulate it into a multi-convex problem which is then solved by the block coordinate descent method, with convergence guarantee by casting it into the block successive upper bound minimization framework. The proposed VOIMU algorithm can yield a stationary-point solution with convergence guarantee, together with some intriguing information of potential outlier pixels though outliers are neither physically modeled in the above problem nor detected in the algorithm operation. Finally, we provide some simulation results and experimental results using real data to demonstrate the efficacy and practical applicability of the proposed VOIMU algorithm.

INDEX TERMS Hyperspectral imaging, endmember variability, outlier effects, block successive upper bound minimization (BSUM), block coordinate descent (BCD) method, alternating direction method of multipliers (ADMM).

I. INTRODUCTION

A hyperspectral sensor records electromagnetic fingerprints (scattering patterns) of substances (materials) in a scene, known as spectral signatures, over hundreds of narrow spectral bands (typically, 5-10 nm in wavelength spacing) from visible to short-wave infrared wavelength region. However, mixed pixel spectra (i.e., mixed pixel phenomena) are prevalent in the hyperspectral image/data due to limited spatial resolution of the hypersepctral sensor, and thus they must be decomposed for identifying the underlying materials for a variety of applications in hyperspectral remote sensing (HRS) [1]–[3] such as planetary exploration, land mapping and classification, environmental monitoring, and mineral identification and quantification [4]–[6]. To accurately identify materials (or endmembers) as well as their corresponding proportions (or abundances) for a scene of interest, hyperspectral unmixing (HU) [7], [8] with the observed hyperspectral data has been extensively studied over the last 2 decades. Indeed many powerful HU algorithms have been proposed, especially when the conventional linear mixing model (LMM) can well apply to the hyperspectral data for heavily-mixed mixtures [9] and ill-conditioned mixtures [10], [11]. However, the conventional LMM may not be very suitable for some hyperspectral scenes for which endmember variability (EV) and outlier effects (OE) are non-negligible, and recently, the HU design has drawn considerable attention for such hyperspectral data.

In many scenarios, the conventional LMM bears inadequate approximation, to some extent, of the actual mixing process, when EV [12]–[17] and OE [18]–[20] are present as two confounding factors that hamper the efficacy of HU algorithms. The EV makes conventional LMM invalid mostly because the measured radiance or reflectance of a material changes according to the geometry and topography of the scene, thus leading to each endmember inappropriately represented by a single spectral signature. Several more factors such as atmospheric effects, intrinsic variability of material, and the variation of a hidden parameter (e.g., concentration of chlorophyll in vegetation) also cause EV in hyperspectral data. Moreover, the OE, similarly, is another obstacle we need to overcome. In general, outlier pixels have quite distinct spectra from the background pixels. However, there are plenty of causes making a background pixel unable to deliver reliable information. For example, any defect in electronics makes it difficult to read or access the generated data via sensors, or a reading of bad pixels resulted from defective electronic apparatus exceeds the operation scope of the sensors. Most of the existing HU algorithms either focus on EV or on OE, but none on both simultaneously as far as we know.

Current studies of HU algorithms that can handle EV mainly comprise (i) the spectral bundles based approaches, (ii) the probability distribution based approaches and (iii) physics-model based approaches. For approaches in (i), each endmember is represented as a set of spectra extracted from a number of randomly selected sub-images [21]. However, such approaches do not explicitly model the variability, sometimes leading to uninterpretable unmixing results, and the assumptions (e.g., the pure pixel assumption) required by the endmember extraction algorithm used may not hold for all the sub-images. For approaches in (ii), the M-band spectral signature of each material is assumed to be a random vector following some independent and identically distributed (i.i.d.) probability distribution (e.g., normal composition model (NCM) [13] and beta compositional model (BCM) [14], and the Gaussian mixture model (GMM) [15]). However, such approaches cannot be practically applied in many scenarios due to limitations of exponentially increasing complexity with number of spectral bands (M), number of materials and number of pixels, and presence of regions of pure pixels [15]. For approaches in (iii), a physics-based model is used for modeling the EV such as Hapke model [22], which may lead to cumbersome HU algorithms, extended linear mixing model (ELMM) [17], which can account for illumination variations but lack flexibility when the endmembers are subject to more complex spectral distortions, and perturbed linear mixing model (PLMM) [16], which allows the endmembers to vary from a pixel to another by modeling the EV as an additive perturbation, thus well interpreting spatial-spectral variabilities.

There are also some efforts devoted to the HU algorithm design with OE taken into account. A straightforward approach is to detect outlier pixels and then remove them from the hyperspectral data followed by the standard HU processing. Various outlier detection algorithms have been proposed such as clustering based approaches [23], Reed-Xiaoli (RX) algorithm [24] that statistically detects outliers using a sliding window, and random-selection-based anomaly detector (RSAD) [19], which is an improved version of RX algorithm. Recently, a robust affine set fitting (RASF) approach proposed in [20], considers the background as the conventional LMM of different endmembers and then finds the best affine set of the given hyperspectral data such that outliers can be separated from background pixels, and thereby effectively detected. However, there are very few effective HU algorithms when both EV and OE are present in the hyperspectral data, thus motivating the study to be presented in this work.

In this paper, we propose a novel robust HU algorithm, referred to as the variability/outlier-insensitive multi-convex unmixing (VOIMU) algorithm, that can handle both EV and OE. Considering two suitable regularizers, a non-convex minimization problem is formulated for which the PLMM is used for modeling EV, while OE is implicitly handled by using a p quasi-norm function for the data fitting error with $0 < p < 1$ [25]–[27]. Then we reformulate it into a multi-convex problem and solve the resulting problem by the block coordinate descent (BCD) [28] method, and then the VOIMU algorithm is designed to obtain the desired solution. Some analyses about the performance and convergence of the proposed VOIMU algorithm are also presented. Finally we provide some simulation results and experimental results using real data to demonstrate the efficacy and practical applicability of the proposed VOIMU algorithm, followed by some conclusions.

The remainder of this paper is organized as follows: Section II presents the signal model and the problem formulation. The proposed VOIMU algorithm is detailed in Section III, including theory and analysis with the associated proofs given in the Appendixes. Then some simulation results and real data experiments are presented in Sections IV and V, respectively. Finally, we conclude the paper in Section VI.

For ease of the ensuing presentation, some notations are defined collectively hereinafter. $\mathbb{R}$ ($\mathbb{R}^N$, $\mathbb{R}^{M\times N}$) is the set of real numbers (N-vectors, $M \times N$ matrices). $\mathbb{R}_+$ ($\mathbb{R}^N_+$, $\mathbb{R}^{M\times N}_+$) is the set of nonnegative real numbers (N-vectors, $M \times N$ matrices). Boldface $\mathbf{x}$ and $\mathbf{X}$ denote a column vector and a matrix, respectively; $\mathbf{0}_{M\times N}$ ($\mathbf{0}_M$), $\mathbf{1}_M$ and $\mathbf{I}_M$, respectively, denote the all-zero $M \times N$ matrix (all-zero M-vector), all-one M-vector, and $M \times M$ identity matrix. $\mathbf{e}_i^{(m)} \in \mathbb{R}^m$ denote the i-th unit vector. $\|\cdot\|_p$, $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the p quasi-norm ($0 < p < 1$), 2-norm and the Frobenius norm, respectively. $\otimes$ and $\succeq$ stand for the Kronecker product and the componentwise inequality, respectively. $[\mathbf{x}]_+$ denotes the vector by replacing all the negative elements in $\mathbf{x}$ with zero. $\mathbf{vec}(\mathbf{X})$ is a column vector formed by sequentially stacking the columns of the matrix $\mathbf{X}$; $\mathbf{devec}(\mathbf{x}, M, N)$ denotes an $M \times N$ matrix for which $\mathbf{x} = \mathbf{vec}(\mathbf{devec}(\mathbf{x}, M, N))$. $\mathcal{I}_L \triangleq \{1, \ldots, L\}$ for any positive integer L. $\{\mathbf{A}_n\}$ represents the set of $\mathbf{A}_n$ for

all the admissible values of n, and $\{\mathbf{B}_n\}_{n=1}^{L} \triangleq \{\mathbf{B}_1, \ldots, \mathbf{B}_L\}$. $\mathbf{DIAG}(\{\mathbf{B}_n\})$ denotes the block diagonal matrix with the n-th diagonal block equal to $\mathbf{B}_n$.

II. SIGNAL MODEL AND PROBLEM FORMULATION

In the absence of any prior information of the EV nature, we adopt the PLMM, for which the data pixel $\mathbf{x}_n$ can be represented as

$$\mathbf{x}_n = \mathbf{A}_n \mathbf{s}_n + \mathbf{w}_n \in \mathbb{R}^M, \quad n \in \mathcal{I}_L,$$

where $\mathbf{A}_n \triangleq [\mathbf{a}_{1,n}, \ldots, \mathbf{a}_{N,n}] \in \mathbb{R}^{M\times N}$ is the associated endmember spectral signature matrix; $\mathbf{s}_n \triangleq [s_{1,n}, \ldots, s_{N,n}]^T \in \mathbb{R}^N$ is the associated abundance vector; $\mathbf{w}_n$ is zero-mean white Gaussian noise; and N, M, and L denote the total numbers of endmembers, spectral bands, and pixels, respectively, where N can be estimated ahead of time such as using information-theoretical based method [29] or signal subspace identification method [30]. Note that in the absence of variability (i.e., $\mathbf{A}_1 = \cdots = \mathbf{A}_L = \mathbf{A}$), the above model reduces to the conventional LMM in the HU [31]. Some standard assumptions pertaining to this model are considered in this work as follows:

(A1) $\mathbf{S} \triangleq [\mathbf{s}_1, \ldots, \mathbf{s}_L] \succeq \mathbf{0}_{N\times L}$ (nonnegativity),
(A2) $\mathbf{S}^T \mathbf{1}_N = \mathbf{1}_L$ (sum-to-one),
(A3) $\mathbf{A}_n \succeq \mathbf{0}_{M\times N}$ (nonnegativity).

With the EV taken into account, the HU problem is to blindly estimate $\{\mathbf{A}_n\}$ and $\mathbf{S}$ from the data matrix $\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_L]$. Let

$$\mathbf{y}_n = \mathbf{x}_n - \mathbf{A}_n \mathbf{s}_n$$

denote the data fitting error vector for the n-th pixel, and $\mathbf{y} = [\|\mathbf{y}_1\|_2, \ldots, \|\mathbf{y}_L\|_2]^T$. A natural data fitting criterion [3] is to minimize

$$\|\mathbf{y}\|_p^p = \sum_{n=1}^{L} \|\mathbf{y}_n\|_2^p \tag{1}$$

with respect to (w.r.t.) $\{\mathbf{A}_n\}$ and $\mathbf{S}$, where $p > 0$. Note that for $p \geq 1$, $\|\mathbf{y}\|_p$ is known as the p-norm, while for $0 < p < 1$, it is so-called p quasi-norm. For instance, it has been widely used for the case of the conventional LMM with $p = 2$, which is however sensitive to outliers. Recently, the p quasi-norm based data fitting criterion [25] by minimizing $\|\mathbf{y}\|_p^p$ (where $0 < p < 1$) has been demonstrated that the OE can be effectively suppressed. To understand how it works, please refer to Figure 1, where a toy example is designed for readers to understand the outlier-insensitivity mechanism of p quasi-norm from a simple optimization perspective.

Example 1: In this example, we illustrate the outlier-insensitivity of p quasi-norm. For illustration purpose (cf. Figure 1), assume no EV and no noise in the data, and consider a dataset $\mathcal{X}$ of $L = 100$ pixels comprising $N = 2$ endmembers, i.e., $\mathcal{X} \triangleq \{\mathbf{x}_n = \mathbf{A}\mathbf{s}_n \mid \mathbf{s}_n = (\frac{n}{L}, 1 - \frac{n}{L}),\ \forall n \in \mathcal{I}_{L-1}\} \cup \{\mathbf{x}_L\}$, where $\mathbf{x}_L \triangleq \frac{1}{2}(\mathbf{a}_1 + \mathbf{a}_2) + \ell\mathbf{c}$ ($\ell > 0$ and $\mathbf{c}$ is any unit-norm vector ($\|\mathbf{c}\|_2 = 1$) orthogonal to $\mathbf{a}_1 - \mathbf{a}_2$) is an outlier not located on the endmember simplex. Such dataset is illustrated in Figure 1. Assuming that the outlier is not too seriously deviated from the endmember simplex (say $\ell < \frac{1}{2}\|\mathbf{a}_1 - \mathbf{a}_2\|_2$), one can verify that $\arg\min_{\mathbf{A}'} \|\mathbf{y}\|_p^p = \mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2] \neq \arg\min_{\mathbf{A}'} \|\mathbf{y}\|_2^2$ for $0 < p < 1$, showing the robust fitting of p quasi-norm over the conventional 2-norm based fitting.

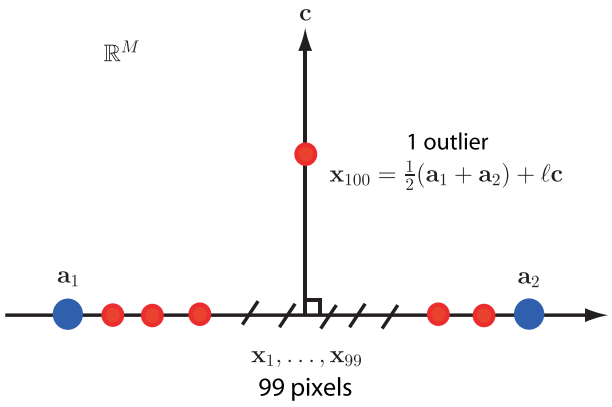

FIGURE 1. Illustration of the dataset $\mathcal{X}$ in Example 1.

In this work, to develop an HU algorithm that is robust against both EV and OE simultaneously, we consider the following nonconvex minimization problem under the assumptions (i.e., constraints) (A1), (A2) and (A3):

$$\begin{aligned} \min_{\{\mathbf{A}_n\},\mathbf{S}} \ & \frac{1}{2}\|\mathbf{y}\|_p^p \\ \text{s.t.} \ & \mathbf{S} \succeq \mathbf{0}_{N\times L},\ \mathbf{S}^T\mathbf{1}_N = \mathbf{1}_L, \\ & \mathbf{A}_n \succeq \mathbf{0}_{M\times N}, \quad \forall n \in \mathcal{I}_L. \end{aligned} \tag{2}$$

Obviously, it is an ill-posed inverse problem, and so we need to use some advisable regularizers in order to promote some desired characteristics of the solution for $\mathbf{A}_n$ for all n and $\bar{\mathbf{A}}$.

Two regularizers are considered as follows:

$$\phi_1(\mathbf{A}_n, \bar{\mathbf{A}}) \triangleq \frac{1}{2}\|\mathbf{A}_n - \bar{\mathbf{A}}\|_F^2,\ \phi_2(\bar{\mathbf{A}}) \triangleq \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\bar{\mathbf{a}}_i - \bar{\mathbf{a}}_j\|_2^2. \tag{3}$$

The first regularizer is to limit EV energy w.r.t. $\bar{\mathbf{A}} \triangleq [\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_N] \in \mathbb{R}^{M\times N}$ (a reference endmember signature matrix), a natural characteristic [16] in HU in the presence of EV. Note that signature matrices at different pixels are supposed to be similar (though different), and minimizing $\phi_1(\mathbf{A}_n, \bar{\mathbf{A}})$ also implicitly promotes such desired solution according to the triangle inequality (i.e., $\|(\mathbf{A}_i - \mathbf{A}_j)\|_F \leq \|\mathbf{A}_i - \bar{\mathbf{A}}\|_F + \|\mathbf{A}_j - \bar{\mathbf{A}}\|_F$, $\forall i, j \in \mathcal{I}_L$). The other regularizer $\phi_2(\bar{\mathbf{A}})$, the so-called sum-of-squared-distances (SSD) regularizer, is to demote the volume of the simplex formed by the reference endmember spectral signatures [32], [33] (i.e., the convex hull of $\{\bar{\mathbf{a}}_i, i = 1, \ldots, N\}$), motivated by the fact that endmember spectral signatures without EV and without noise can be perfectly identified under some mild condition [34]–[36] from the minimum volume enclosing simplex.

Based on (2) together with the above two regularizers taken into account, the proposed robust VOIMU algorithm tries to solve the following nonconvex problem:

$$\begin{aligned}\min_{\{\mathbf{A}_n\},\bar{\mathbf{A}},\mathbf{S}}\ & \frac{1}{2}\sum_{n=1}^{L}\Big[\|\mathbf{x}_n-\mathbf{A}_n\mathbf{s}_n\|_2^2+\epsilon\Big]^{p/2}\\ &+\lambda_1\sum_{n=1}^{L}\phi_1(\mathbf{A}_n,\bar{\mathbf{A}})+\lambda_2\phi_2(\bar{\mathbf{A}}) \qquad (4)\\ \text{s.t. }\ & \mathbf{S}\succeq\mathbf{0}_{N\times L},\ \mathbf{S}^T\mathbf{1}_N=\mathbf{1}_L,\\ & \mathbf{A}_n\succeq\mathbf{0}_{M\times N},\quad \forall n\in\mathcal{I}_L,\end{aligned}$$

where $\lambda_1,\lambda_2 > 0$ are regularization parameters. It is noticeable that the first term in the objective function in problem (4) is obtained by modifying $\|\mathbf{y}_n\|_2^p$ in (1) as $(\|\mathbf{y}_n\|_2^2+\epsilon)^{p/2}$ (in which ϵ is a small positive real number) to avoid numerical problems in the algorithm design in Section III (cf. Remark 1).

III. THE PROPOSED VOIMU ALGORITHM

In this section, we propose an algorithm to solve the nonconvex problem (4). Note that the feasible set of (4) is convex and the two regularization terms in the objective function are also convex, but the only term involving p quasi-norm is nonconvex and also the bottleneck to reformulate (4) into a tractable problem.

The idea is to formulate the nonconvex p quasi-norm term in (4) into a multi-convex function and then the resulting objective function will also be a multi-convex function. Suppose that $\mathcal{C}\subset\mathbb{R}$ is a nonempty convex set, and define a function $f_\varepsilon:\mathbb{R}^2\longrightarrow\mathbb{R}$ in the following form

$$f_\varepsilon(v,z)\triangleq g_\varepsilon(z)v^2+h_\varepsilon(z), \qquad (5)$$

where $g_\varepsilon(z):\mathbb{R}\longrightarrow\mathbb{R}$, $h_\varepsilon(z):\mathbb{R}\longrightarrow\mathbb{R}$, and $\varepsilon>0$ is a parameter. To adopt BCD in solving the nonconvex problem, we have to properly choose $g_\varepsilon(z)$ and $h_\varepsilon(z)$ such that $g_\varepsilon(\mathcal{C})\subset\mathbb{R}_+$ (implying that f_ε is convex in v), $f_\varepsilon(v,z)$ is convex in z, and

$$(v^2+\varepsilon)^{p/2}=\min_{z\in\mathcal{C}}f_\varepsilon(v,z). \qquad (6)$$

How to choose $g_\varepsilon(z)$, $h_\varepsilon(z)$ will be presented in Lemma 1 in Subsection III-A.

By applying (5) and (6), in which $v=\|\mathbf{x}_n-\mathbf{A}_n\mathbf{s}_n\|_2$, $z=z_n$ and $\varepsilon=\epsilon$, to the p quasi-norm term of (4), we come up with the following multi-convex optimization problem:

$$\begin{aligned}\min_{\mathbf{z},\{\mathbf{A}_n\},\bar{\mathbf{A}},\mathbf{S}}\ & \frac{1}{2}\sum_{n=1}^{L}\big(g_\epsilon(z_n)\|\mathbf{x}_n-\mathbf{A}_n\mathbf{s}_n\|_2^2+h_\epsilon(z_n)\big)\\ &+\lambda_1\sum_{n=1}^{L}\phi_1(\mathbf{A}_n,\bar{\mathbf{A}})+\lambda_2\phi_2(\bar{\mathbf{A}}) \qquad (7)\\ \text{s.t. }\ & \mathbf{S}\succeq\mathbf{0}_{N\times L},\ \mathbf{S}^T\mathbf{1}_N=\mathbf{1}_L,\ \ \mathbf{z}\in\mathcal{C}^L,\\ & \mathbf{A}_n\succeq\mathbf{0}_{M\times N},\quad \forall n\in\mathcal{I}_L,\end{aligned}$$

where $\mathbf{z}\triangleq[z_1,\ldots,z_L]^T$ is an auxiliary variable and $\mathcal{C}$ is a preassigned nonempty closed convex subset of $\mathbb{R}$. Then the BCD method [28] can be applied to solve problem (7) to be presented next.

For ease of the ensuing presentation, let

$$\mathcal{J}(\mathbf{z},\{\mathbf{A}_n\},\mathbf{S})=\frac{1}{2}\sum_{n=1}^{L}\big(g_\epsilon(z_n)\|\mathbf{x}_n-\mathbf{A}_n\mathbf{s}_n\|_2^2+h_\epsilon(z_n)\big). \qquad (8)$$

The BCD method is to alternatively update each unknown variables of $\mathbf{z}$, $\{\mathbf{A}_n\}$, $\bar{\mathbf{A}}$, and $\mathbf{S}$ in a round-robin manner by solving the associated convex subproblems as follows:

$$\mathbf{z}^{k+1}\in\arg\min_{\mathbf{z}\in\mathcal{C}^L}\mathcal{J}(\mathbf{z},\{\mathbf{A}_n^k\},\mathbf{S}^k), \qquad (9)$$

$$\begin{aligned}\{\mathbf{A}_n^{k+1}\}\in\arg\min_{\{\mathbf{A}_n\}}\ & \mathcal{J}(\mathbf{z}^{k+1},\{\mathbf{A}_n\},\mathbf{S}^k)+\lambda_1\sum_{n=1}^{L}\phi_1(\mathbf{A}_n,\bar{\mathbf{A}}^k) \qquad (10)\\ \text{s.t. }\ & \mathbf{A}_n\succeq\mathbf{0}_{M\times N},\quad \forall n\in\mathcal{I}_L,\end{aligned}$$

$$\bar{\mathbf{A}}^{k+1}\in\arg\min_{\bar{\mathbf{A}}}\lambda_1\sum_{n=1}^{L}\phi_1(\mathbf{A}_n^{k+1},\bar{\mathbf{A}})+\lambda_2\phi_2(\bar{\mathbf{A}}), \qquad (11)$$

$$\begin{aligned}\mathbf{S}^{k+1}\in\arg\min_{\mathbf{S}}\ & \mathcal{J}(\mathbf{z}^{k+1},\{\mathbf{A}_n^{k+1}\},\mathbf{S}) \qquad (12)\\ \text{s.t. }\ & \mathbf{S}\succeq\mathbf{0}_{N\times L},\quad \mathbf{S}^T\mathbf{1}_N=\mathbf{1}_L,\end{aligned}$$

where the superscript "k" denotes the iteration number, and $\{\mathbf{A}_n^0\}$, $\bar{\mathbf{A}}^0$ and $\mathbf{S}^0$ can be initialized by successive decoupled volume max-min (SDVMM)-RASF/fully constrained least squares (FCLS) [20], [37], which is one of the state-of-the-art HU algorithms against OE. Solving (9), (10), (11) and (12), will be presented in the ensuing subsections, respectively. The resulting BCD algorithm is summarized in Algorithm 1, and it can be shown that Algorithm 1 will converge and yield a stationary-point solution to problem (7), as stated in the following proposition with the proof given in Appendix A.

Proposition 1: Assume that all the convex subproblems (9), (10), (11), *and* (12) *can be optimally solved at each iteration. Then the sequence* $\{\mathbf{z}^k,\{\mathbf{A}_n^k\},\bar{\mathbf{A}}^k,\mathbf{S}^k\}$ *generated by Algorithm 1 converges to a stationary point of* (7).

Though $\mathbf{z}^{k+1}$ and $\bar{\mathbf{A}}^{k+1}$ can be updated with the associated closed-form expressions, the convex problems (10) and (12) are large-scale convex problems, so they cannot be efficiently solved using the off-the-shelf convex solvers (e.g., CVX [38]) for updating $\{\mathbf{A}_n^{k+1}\}$ and $\mathbf{S}^{k+1}$, respectively. The alternating direction method of multipliers (ADMM) [39] has been recognized as an effective method for solving large-scale convex optimization problem. Hence, we employ it to solve (10), whereas problem (12) is actually a fully constrained weighted least-squares problem which can be efficiently solved by the existing FCLS algorithm [37]. The details will be presented in the following subsections.

Algorithm 1 BCD Algorithm for Solving (7)

1: **Given X**, and initial $\bar{\mathbf{A}}^0$ and $\mathbf{S}^0$ obtained by SDVMM-RASF/FCLS [20]
2: Set $k := 0$ and $\mathbf{A}_n^0 := \bar{\mathbf{A}}^0,\ \forall n \in \mathcal{I}_L$.
3: **repeat**
4: Update $\mathbf{z}^{k+1}$ using the closed-form solution (15);
5: Update $\{\mathbf{A}_n^{k+1}\}$ by solving (10) using Algorithm 2;
6: Update $\bar{\mathbf{A}}^{k+1}$ using the closed-form solution (26);
7: Update $\mathbf{S}^{k+1}$ by solving (12) using FCLS Algorithm [37];
8: $k := k + 1$;
9: **until** pre-defined stopping criterion is satisfied.
10: **Output** $\mathbf{z}^k$, $\{\mathbf{A}_n^k\}$, $\bar{\mathbf{A}}^k$, $\mathbf{S}^k$.

A. CLOSED-FORM SOLUTION FOR SOLVING (9)

To ensure that the function values of $g_\varepsilon(z)$ is non-negative (thus guaranteeing the convexity of $f_\varepsilon(v, z)$ w.r.t. v; cf. (5), (6)), we choose $g_\varepsilon(z) = z^2$. In the following Lemma, proved in Appendix B, we show that the counterpart function $h_\varepsilon(z)$ (for such $g_\varepsilon(z)$) also exists for (6) to be satisfied.

Lemma 1: Assume that $0 < p < 1$ *and* $\varepsilon > 0$. *Let* $\mathcal{C} = \mathbb{R}_+$, $\alpha = (2/p)^{\frac{p}{p-2}} - (2/p)^{\frac{2}{p-2}} > 0$, *and*

$$g_\varepsilon(z) \triangleq z^2,\ h_\varepsilon(z) \triangleq \alpha z^{\frac{2p}{p-2}} + \varepsilon z^2 \tag{13}$$

for which $g_\varepsilon(\mathcal{C}) \subset \mathbb{R}_+$. Then $f_\varepsilon(v, z)$ defined in (5) is convex in z over $\mathcal{C}$,

$$z^\star = \arg\min_{z\in\mathcal{C}} f_\varepsilon(v, z) = \left[\frac{2-p}{\alpha p}\left(v^2+\varepsilon\right)\right]^{\frac{p-2}{4}}, \tag{14}$$

and the optimal value $f_\varepsilon(v, z^\star)$ *is given by* (6).

It is noticeable from (8), that problem (9) can actually be solved in a decoupling manner (i.e., pixel-wise manner). By setting $v = \|\mathbf{x}_n - \mathbf{A}_n^k\mathbf{s}_n^k\|_2$ and $\varepsilon = \epsilon$ in (14), it can be readily inferred that the optimal solution z_n^{k+1} to convex problem (9) with $\mathcal{C} = \mathbb{R}_+$, $g_\epsilon(z_n)$ and $h_\epsilon(z_n)$ according to (13), is obtained as

$$z_n^{k+1} = \left[\frac{2-p}{\alpha p}\left(\|\mathbf{x}_n - \mathbf{A}_n^k\mathbf{s}_n^k\|_2^2+\epsilon\right)\right]^{\frac{p-2}{4}}, \quad \forall n \in \mathcal{I}_L, \tag{15}$$

where the parameter α is given in Lemma 1.

Remark 1: In spite of the closed-form optimal solution $z_n^\star$ given by (15), if $\epsilon = 0$ in (15), computing the optimal $z_n^\star$ may lead to numerical problems when the data fitting term $\|\mathbf{x}_n - \mathbf{A}_n^k\mathbf{s}_n^k\|_2$ is significantly smaller than unity such that $z_n^\star$ may get almost unbounded for some n because of the negative exponent $-1/2 < (p-2)/4 < -1/4$. Meanwhile, when $\|\mathbf{x}_n - \mathbf{A}_n^k\mathbf{s}_n^k\|_2 \ll 1$, the larger the ϵ (where $\epsilon < 1$), the smaller the $z_n^\star$. Hence it is advisable to choose a moderate value for ϵ (e.g., $\epsilon = 10^{-3}$) such that it will not cause the neglect of the two regualarizers in (7).

B. ALGORITHM FOR SOLVING (10)

By letting $g_\epsilon(z_n^{k+1}) = (z_n^{k+1})^2$ (cf. (13)) in (10), solving problem (10) is equivalent to solve the following problem:

$$\min_{\{\mathbf{A}_n\}} \frac{1}{2}\sum_{n=1}^{L}\|z_n^{k+1}\mathbf{x}_n - z_n^{k+1}\mathbf{A}_n\mathbf{s}_n^k\|_2^2 + \frac{\lambda_1}{2}\sum_{n=1}^{L}\|\mathbf{A}_n - \bar{\mathbf{A}}^k\|_F^2 \tag{16}$$
$$\text{s.t. } \mathbf{A}_n \succeq \mathbf{0}_{M\times N}, \quad \forall n \in \mathcal{I}_L.$$

Apparently, (16) can also be solved in a pixel-wise manner. Instead, we solve this problem in block-wise manner due to no closed-form solution for $\mathbf{A}_n$, in order to have better computational efficiency, especially when L is large.

Suppose that the j-th data block consists of $\mathbf{x}_n, n = (j-1)L_s + 1, \ldots, jL_s$, where $j = 1, \ldots, \lceil L/L_s \rceil$ (the smallest integer larger than or equal to L/L_s, implying that the last block size is less than L_s if L/L_s is not an integer). The corresponding j-th subproblem, through some change of variables, can be reformulated as the following unconstrained convex problem

$$\min_{\mathbf{c}_j} \frac{1}{2}\|\mathbf{\Phi}_j\mathbf{c}_j - \mathbf{q}_j\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{c}_j - \bar{\mathbf{a}}\|_2^2 + I_+(\mathbf{c}_j), \tag{17}$$

where

$$\mathbf{c}_j = \mathbf{vec}([\mathbf{A}_{(j-1)L_s+1}, \ldots, \mathbf{A}_{jL_s}]) \in \mathbb{R}^{MNL_s}, \tag{18a}$$
$$\mathbf{q}_j = \left[z_{(j-1)L_s+1}^{k+1}\mathbf{x}_{(j-1)L_s+1}^T, \ldots, z_{jL_s}^{k+1}\mathbf{x}_{jL_s}^T\right]^T, \tag{18b}$$
$$\mathbf{\Phi}_j = \mathbf{DIAG}\left(\{(z_n^{k+1}\mathbf{s}_n^k)^T\}_{n=(j-1)L_s+1}^{jL_s} \otimes \mathbf{I}_M\right), \tag{18c}$$
$$\bar{\mathbf{a}} = \mathbf{1}_{L_s} \otimes \mathbf{vec}(\bar{\mathbf{A}}^k) \in \mathbb{R}^{MNL_s}, \tag{18d}$$
$$I_+(\mathbf{c}_j) \triangleq \begin{cases} 0, & \mathbf{c}_j \succeq \mathbf{0}_{MNL_s} \\ \infty, & \text{otherwise.} \end{cases} \tag{18e}$$

To solve (17) using ADMM [39], subproblem (17) needs to be re-expressed as the following form:

$$\min_{\mathbf{c}_j,\mathbf{t}} \frac{1}{2}\|\mathbf{\Phi}_j\mathbf{c}_j - \mathbf{q}_j\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{c}_j - \bar{\mathbf{a}}\|_2^2 + I_+(\mathbf{t}) \tag{19}$$
$$\text{s.t. } \mathbf{c}_j = \mathbf{t}.$$

Then the augmented Lagrangian of (19) is given by

$$\mathcal{L}(\mathbf{c}_j, \mathbf{t}, \mathbf{d}) = \frac{1}{2}\|\mathbf{\Phi}_j\mathbf{c}_j - \mathbf{q}_j\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{c}_j - \bar{\mathbf{a}}\|_2^2 + I_+(\mathbf{t}) + \frac{\eta}{2}\|\mathbf{c}_j - \mathbf{t} + \mathbf{d}\|_2^2,$$

where $\mathbf{d}$ is the scaled dual variable and $\eta > 0$ is penalty parameter. The ADMM solves (19) by alternatively updating two primal variables and the dual variable as follows:

$$\mathbf{t}^{i+1} \in \arg\min_{\mathbf{t}} \mathcal{L}(\mathbf{c}_j^i, \mathbf{t}, \mathbf{d}^i), \tag{20a}$$
$$\mathbf{c}_j^{i+1} \in \arg\min_{\mathbf{c}_j} \mathcal{L}(\mathbf{c}_j, \mathbf{t}^{i+1}, \mathbf{d}^i), \tag{20b}$$
$$\mathbf{d}^{i+1} = \mathbf{d}^i + \mathbf{c}_j^{i+1} - \mathbf{t}^{i+1}, \tag{20c}$$

where $\mathbf{c}_j^0$ and $\mathbf{d}^0$ are initialized by $\mathbf{0}_{MNL_s}$ (or by warm start [39]), and the superscript "i" denotes the iteration number of ADMM algorithm.

Problem (20a) is generally referred to as the proximity operator for the indicator function $I_+(\mathbf{t})$ [39], and its solution is known as

$$\mathbf{t}^{i+1} = [\mathbf{c}_j^i + \mathbf{d}^i]_+. \tag{21}$$

Moreover, the closed-form solution to the unconstrained quadratic convex problem (20b) can be easily shown to be

$$\mathbf{c}_j^{i+1} = (\mathbf{\Phi}_j^T \mathbf{\Phi}_j + (\lambda_1 + \eta)\mathbf{I}_{MNL_s})^{-1}(\mathbf{\Phi}_j^T \mathbf{q}_j + \boldsymbol{\nu}), \tag{22}$$

where $\boldsymbol{\nu} = \lambda_1 \bar{\mathbf{a}} + \eta \mathbf{t}^{i+1} - \eta \mathbf{d}^i$. The resulting ADMM algorithm is summarized in Algorithm 2, and meanwhile, convergence can also be guaranteed because the convergence condition w.r.t. the equality constraint in problem (19) is satisfied [39].

Note that based on our experience, as $L_s = L$ (only one data block), computing $\mathbf{c}_j^{i+1}$ (cf. (22)) using available software such as MATLAB could yield memory overflow problems in $\mathbf{\Phi}_j$ (cf. (18c)) even when L is moderately large. Moreover, as $L_s = 1$, we actually need to solve the problem (19) using ADMM L times, so computationally too costly to be practical when L is large. Hence, the preceding multi-block processing turns out to be a feasible solution. However, a large-size ($MNL_s \times MNL_s$) matrix inverse is involved in updating $\mathbf{c}_j^{i+1}$ (cf. (22)), thus yielding high computational cost as well. Instead, an alternative expression for $\mathbf{c}_j^{i+1}$, which is computationally much more efficient than (22), is proposed in the following proposition.

Proposition 2: The equation (22) for computing $\mathbf{c}_j^{i+1}$ is identical to

$$\mathbf{c}_j^{i+1} = \mathbf{vec}\Big(\big(\mathbf{devec}(\mathbf{q}_j, M, L_s)\mathbf{\Lambda} + \mathbf{\Gamma}\big)\mathbf{\Psi}\Big), \tag{23}$$

where

$$\mathbf{\Psi} = \mathbf{DIAG}\Big(\Big\{\big((z_n^{k+1})^2 \mathbf{s}_n^k (\mathbf{s}_n^k)^T + (\lambda_1 + \eta)\mathbf{I}_N\big)^{-1}\Big\}_{n=l}^{jL_s}\Big), \tag{24a}$$

$$\mathbf{\Lambda} = \mathbf{DIAG}\big(\{(z_n^{k+1}\mathbf{s}_n^k)^T\}_{n=l}^{jL_s}\big) \in \mathbb{R}^{L_s \times NL_s}, \tag{24b}$$

$$\mathbf{\Gamma} = \mathbf{devec}(\boldsymbol{\nu}, M, NL_s), \tag{24c}$$

and the parameter $l \triangleq (j-1)L_s + 1$ in (24a) and (24b).

The proof of Proposition 2 is relegated to Appendix C. Note that the matrix inversion in the closed-form solution of Proposition 2 is performed only for a matrix size of $N \times N$ (cf. (24a)), where N is usually within ten. Moreover, the multi-block processing can be implemented by MATLAB Parallel Computing Toolbox for better computational efficiency.

C. CLOSED-FORM SOLUTION FOR SOLVING (11)

By substituting ϕ_1 defined in (3) into problem (11), it can be rewritten in the following form:

$$\bar{\mathbf{A}}^{k+1} = \arg\min_{\bar{\mathbf{A}}} \frac{\lambda_1}{2} \sum_{n=1}^{L} \|\mathbf{A}_n^{k+1} - \bar{\mathbf{A}}\|_F^2 + \lambda_2 \phi_2(\bar{\mathbf{A}}) \tag{25}$$

where $\phi_2(\bar{\mathbf{A}})$ also defined in (3) can be re-expressed as

$$\phi_2(\bar{\mathbf{A}}) = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \|\mathbf{P}_{ij}\bar{\mathbf{a}}\|_2^2 = \frac{1}{2}\|\mathbf{P}\bar{\mathbf{a}}\|_2^2,$$

Algorithm 2 ADMM for Solving (10)

1: **Given** $\mathbf{X}$, $\mathbf{z}^{k+1}$, $\bar{\mathbf{A}}^k$, $\mathbf{S}^k$ and L_s.
2: Set $j := 1$.
3: **repeat**
4: Initialize $\mathbf{c}_j^0$, $\mathbf{d}^0 = \mathbf{0}_{MNL_s}$.
5: Set $i := 0$.
6: **repeat**
7: Update $\mathbf{t}^{i+1}$ by (21);
8: Update $\mathbf{c}_j^{i+1}$ by (23);
9: Update $\mathbf{d}^{i+1} := \mathbf{d}^i + \mathbf{c}_j^{i+1} - \mathbf{t}^{i+1}$;
10: $i := i + 1$;
11: **until** pre-defined stopping criterion is satisfied.
12: $j := j + 1$;
13: **until** $j = \lceil L/L_s \rceil + 1$.
14: **Output** $\{\mathbf{A}_n^{k+1}\}$.

where $\bar{\mathbf{a}} \triangleq \mathbf{vec}(\bar{\mathbf{A}})$, $\mathbf{P}_{ij} \triangleq (\mathbf{e}_i^{(N)} - \mathbf{e}_j^{(N)})^T \otimes \mathbf{I}_M$ and $\mathbf{P} \in \mathbb{R}^{0.5MN(N-1)\times MN}$ is the matrix formed by stacking all the $\mathbf{P}_{ij}$. By letting $\mathbf{a}_n^{k+1} = \mathbf{vec}(\mathbf{A}_n^{k+1})$, $\forall n \in \mathcal{I}_L$, problem (25) is an unconstrained quadratic convex problem, and its solution can be easily obtained as

$$\bar{\mathbf{a}}^{k+1} = (\lambda_2 \mathbf{P}^T\mathbf{P} + L\lambda_1 \mathbf{I}_{MN})^{-1}(\lambda_1 \sum_{n=1}^{L} \mathbf{a}_n^{k+1}). \tag{26}$$

The optimal solution for $\bar{\mathbf{A}}^{k+1} = \mathbf{devec}(\bar{\mathbf{a}}^{k+1}, M, N)$ can be obtained from (26).

D. ALGORITHM FOR SOLVING (12)

Due to $g_\epsilon(z_n^{k+1}) = (z_n^{k+1})^2$ (cf. (13)), problem (12) can be reformulated as

$$\min_{\mathbf{S}} \ \frac{1}{2}\sum_{n=1}^{L}(z_n^{k+1})^2 \|\mathbf{x}_n - \mathbf{A}_n^{k+1}\mathbf{s}_n\|_2^2$$
$$\text{s.t. } \mathbf{S} \succeq \mathbf{0}_{N\times L},\ \mathbf{S}^T\mathbf{1}_N = \mathbf{1}_L.$$

This problem is nothing but a weighted least-squares convex problem with nonnegativity and sum-to-one constraints, and can be decomposed into L FCLS problems [37]. The resulting solution can be approximately but much more efficiently obtained by using nonnegativity constrained least squares (NCLS) method [37] as follows:

$$\mathbf{s}_n^{k+1} = \arg\min_{\mathbf{s}_n} \left\| \begin{bmatrix} \delta(z_n^{k+1}\mathbf{x}_n) \\ 1 \end{bmatrix} - \begin{bmatrix} \delta(z_n^{k+1}\mathbf{A}_n^{k+1}) \\ \mathbf{1}_N^T \end{bmatrix} \mathbf{s}_n \right\|_2^2,$$
$$\text{s.t. } \mathbf{s}_n \succeq \mathbf{0}_N,$$

where $\delta > 0$ is a small parameter for controlling the impact of the sum-to-one constraint.

Let us conclude this section with following three remarks:

Remark 2: Regarding the stopping criteria for the proposed Algorithm 1 for solving (7) and Algorithm 2 for solving (10), the one used for the former can be relative change of the consecutive objective values smaller than a suitable tolerance (e.g., 10^{-3}), a commonly used stopping

criterion in BCD algorithms [28]. The one used for the latter in each block processing can be the associated primal residual $\|\mathbf{c}_j^i - \mathbf{t}^i\|_2$ and dual residual $\| - \eta(\mathbf{t}^i - \mathbf{t}^{i-1})\|_2$ (where j is the block number) are smaller than a suitable tolerance (e.g., 10^{-3}), also a standard stopping criterion in ADMM [39].

Remark 3: It can be observed from (15) that, if $\mathbf{x}_n$ is an outlier, then $u = \|\mathbf{x}_n - \mathbf{A}_n^k \mathbf{s}_n^k\|_2^2 + \epsilon \gg 1$, and thus $z_n^{k+1} = [(2-p)u/(\alpha p)]^{(p-2)/4} \ll 1$, since $(p-2)/4 < 0$ and $(2-p)/(\alpha p) > 0$. Hence, the pixel n could be an outlier if $1/z_n^{k+1} \gg 1$. Nevertheless, the detection of outliers is not the purpose for the proposed VOIMU algorithm. Furthermore, from (16), it can be inferred that the obtained solution $\mathbf{A}_n^{k+1} \approx \bar{\mathbf{A}}^k$ when $z_n^{k+1} \ll 1$.

Remark 4: From (3) and (4), one can observe that if the m-th endmember does not contribute to the n-th pixel, then the EV energy of m-th endmember must be zero. This observation implies that EV will share the same distribution pattern as the associated abundance map and which can be justified in Sections IV and V (cf. Figures 3, 5, 10, and 16).

IV. SIMULATION RESULTS

This section evaluates the performance of the proposed VOIMU algorithm (Algorithm 1) using synthetic data. As far as we know, no existing HU algorithms are reported in the literature that can handle the EV and OE simultaneously. So, we compare Algorithm 1 with SDVMM-RASF/FCLS algorithm [20], [37] that is only robust against OE, and PLMM [16] and ELMM [17] algorithms that can only handle EV, and vertex component analysis (VCA)/FCLS [37], [40] that is a benchmark HU algorithm based on the conventional LMM without considering both EV and OE, just serving as a baseline for performance comparison. Next, we present generation of synthetic data, performance measures used, and performance comparison in the ensuing subsections, respectively.

A. GENERATION OF SYNTHETIC DATA

Each synthetic dataset is generated with $N = 6$ endmember spectral signatures of $M = 200$ spectral bands, including Alunite, Buddingtonite, Calcite, Jarosite, Muscovite, and Andradite, randomly selected from U.S. geological survey (USGS) library [41], to form the endmember signature matrix

$$\mathbf{A}_{true} = [\mathbf{a}_{1_{true}}, \ldots, \mathbf{a}_{N_{true}}] \in \mathbb{R}^{M \times N}. \tag{27}$$

Then the associated N abundance maps (each being a row vector of the abundance matrix $\mathbf{S}$) taken from [42] that satisfy the assumptions (A1) and (A2) are used to obtain a dataset of size $L = 100 \times 100$ through the following procedure:

(S1) Generate

$$\mathbf{A}_n = [\mathbf{a}_{1_{true}} + \mathbf{p}_{1n}, \ldots, \mathbf{a}_{N_{true}} + \mathbf{p}_{Nn}], \quad \forall n \in \mathcal{I}_L, \tag{28}$$

where $\mathbf{p}_{in}$ is a Gaussian random vector with zero mean and squared-exponential covariance matrix $10^{-3} \times \mathbf{\Sigma}$ [43] with the (i,j)-th element given by

$$[\mathbf{\Sigma}]_{ij} = \exp\left(-(i-j)^2/(M/2)^2\right), \ i,j \in \mathcal{I}_M.$$

(S2) Compute

$$\mathbf{x}_{n_{true}} \triangleq \mathbf{A}_n \mathbf{s}_n, \ \forall n \in \mathcal{I}_L. \tag{29}$$

(S3) Obtain $\mathbf{x}_n$ by adding zero-mean white Gaussian noise to $\mathbf{x}_{n_{true}}$ for a specified signal-to-noise ratio (SNR$= \sum_{n=1}^{L} \|\mathbf{x}_{n_{true}}\|_2^2/(\sigma^2 ML)$ where σ^2 is the noise variance).

(S4) Following the same way as in [20], generate Z outliers $(\mathbf{r}_{l_i}, \ \forall i = 1, \ldots, Z)$ with the outlier pixel indices set $\mathcal{Z} \triangleq \{l_1, \ldots, l_Z\}$, with l_i randomly selected from $\mathcal{I}_L$, as follows:

$$\mathbf{r}_{l_i} = c\boldsymbol{\kappa}_i, \ i = 1, \ldots, Z, \tag{30}$$

where $\boldsymbol{\kappa}_i$ is a zero-mean unit-variance Laplace random vector and c is the scalar to meet a specified signal-to-outlier-ratio (SOR) defined as

$$\text{SOR} = \Big(\sum_{n=1}^{L} \|\mathbf{x}_n\|_2^2/L\Big)/\Big(\sum_{i=1}^{Z} \|\mathbf{r}_{l_i}\|_2^2/Z\Big).$$

Finally, add $\mathbf{r}_{l_i}$ to $\mathbf{x}_{l_i}$ for all $l_i \in \mathcal{Z}$.

B. PERFORMANCE MEASURES

Five commonly used performance measures are as follows:

- Average spectral signature root mean square error (aRMSE) [17]:

$$a\text{RMSE} \triangleq \frac{1}{L-Z} \sum_{n \in \mathcal{I}_L \setminus \mathcal{Z}} \sqrt{\frac{1}{MN} \|\mathbf{A}_n - \widehat{\mathbf{A}}_n\|_F^2},$$

  where $\mathbf{A}_n$ is given by (28) and $\widehat{\mathbf{A}}_n$ is the associated endmember matrix estimate.

- Spectral angle error (SAE) [31]:

$$\text{SAE} \triangleq \sqrt{\frac{1}{N} \sum_{m=1}^{N} \left[\arccos \frac{(\mathbf{a}_{m_{true}})^T \bar{\mathbf{a}}_m}{\|\mathbf{a}_{m_{true}}\|_2 \cdot \|\bar{\mathbf{a}}_m\|_2}\right]^2},$$

  where $\mathbf{a}_{m_{true}}$ is the m-th column of $\mathbf{A}_{true}$ (cf. (27)), and $\bar{\mathbf{a}}_m$ is the associated reference endmember signature estimate.

- Abundance angle error (AAE) [31]:

$$\text{AAE} \triangleq \sqrt{\frac{1}{N} \sum_{m=1}^{N} \left[\arccos \frac{(\boldsymbol{s}_m)^T \widehat{\boldsymbol{s}}_m}{\|\boldsymbol{s}_m\|_2 \cdot \|\widehat{\boldsymbol{s}}_m\|_2}\right]^2},$$

  where $\boldsymbol{s}_m \triangleq [s_{m,1}, \ldots, s_{m,L}]^T \in \mathbb{R}^L$ is the m-th row of $\mathbf{S}$, i.e., the abundance map of the m-th material, and $\widehat{\boldsymbol{s}}_m$ is the associated estimate.

- Average reconstruction spectral angle mapper (xSAM) [17]:

$$x\text{SAM} \triangleq \frac{1}{L-Z} \sum_{n \in \mathcal{I}_L \setminus \mathcal{Z}} \arccos\left(\frac{(\mathbf{x}_{n_{true}})^T \widehat{\mathbf{x}}_n}{\|\mathbf{x}_{n_{true}}\|_2 \cdot \|\widehat{\mathbf{x}}_n\|_2}\right),$$

  where $\mathbf{x}_{n_{true}}$ is given by (29) and $\widehat{\mathbf{x}}_n = \widehat{\mathbf{A}}_n \widehat{\mathbf{s}}_n$ is the associated estimate.

- Reconstruction error (RE) [16]:

$$\text{RE} \triangleq \frac{1}{M(L-Z)} \sum_{n \in \mathcal{I}_L \setminus \mathcal{Z}} \|\mathbf{x}_{n_{true}} - \widehat{\mathbf{x}}_n\|_2^2.$$

Note that $\mathbf{x}_{n_{true}}$ in the performance measures xSAM and RE need to be replaced by measurements $\mathbf{x}_n$ in the real data experiments below due to lack of $\mathbf{x}_{n_{true}}$.

C. PERFORMANCE COMPARISON

Thirty synthetic datasets were generated under the following two scenarios:

- Scenario 1 (with EV, without OE): The data were generated through Steps (S1)-(S3) in Subsection IV-A ($Z = 0$ and $\mathcal{Z}$ is an empty set).
- Scenario 2 (with both EV and OE): The data were generated through Steps (S1)-(S4) in Subsection IV-A ($Z = 10$ and SOR $= -10$ dB).

Then the generated data were processed by all the algorithms under test, and the preceding performance measures and the running time T were calculated by averaging over the performed 30 independent runs. The parameters used for the proposed VOIMU algorithm are $p = 0.5$, $\lambda_1 = 0.5$, $\lambda_2 = 10$, $\epsilon = 10^{-3}$, and $L_s = 25$. For performance comparison, the proposed VOIMU algorithm and the existing VCA/FCLS, SDVMM-RASF/FCLS, PLMM (with parameters $(\alpha, \beta, \gamma) = (10^{-5}, 4.9 \times 10^{-3}, 1)$) and ELMM (with parameters $(\lambda_S, \lambda_A, \lambda_\psi) = (6.3 \times 10^{-2}, 5 \times 10^{-3}, 9 \times 10^{-1})$) algorithms are tested with the same synthetic data. The source codes of the SDVMM-RASF/FCLS [20], PLMM [16] and ELMM [17] algorithms were downloaded from the associated authors' websites. Based on the reference parameters provided in [16] and [17], we also tried our best in finding the proper parameters for their best performances. The simulation results for the two scenarios are shown in Table 1, that were obtained using Mathworks MATLAB R2017b running on a personal computer equipped with Core-i7-4790K CPU with 4-GHz speed and 16-GB random access memory.

It can be observed from Table 1 for Scenario 1, that in terms of RE and xSAM (data-fitting error based performance measures), the VOIMU algorithm significantly outperforms all the other algorithms. ELMM performs much better than PLMM, and PLMM performs slightly better than VCA/FCLS and SDVMM-RASF/FCLS (due to EV never considered in their design). However, all the algorithms under test are comparable in terms of aRMSE, SAE (endmember spectral signature estimation accuracies), and AAE (abundance estimation accuracies), indicating that the estimated reference spectral signatures are also good approximations to $\mathbf{A}_{true}$ "the ground truth signatures" (cf. (27)), and that the estimated abundance maps are also quite similar as shown in Figure 2 for a typical realization. Moreover, for the VOIMU algorithm, the spatial distribution of the estimated EV in terms of

$$\frac{1}{\sqrt{M}} \|\mathbf{a}_{m,n} - \bar{\mathbf{a}}_m\|_2 \text{ (i.e., square root of EV energy)} \quad (31)$$

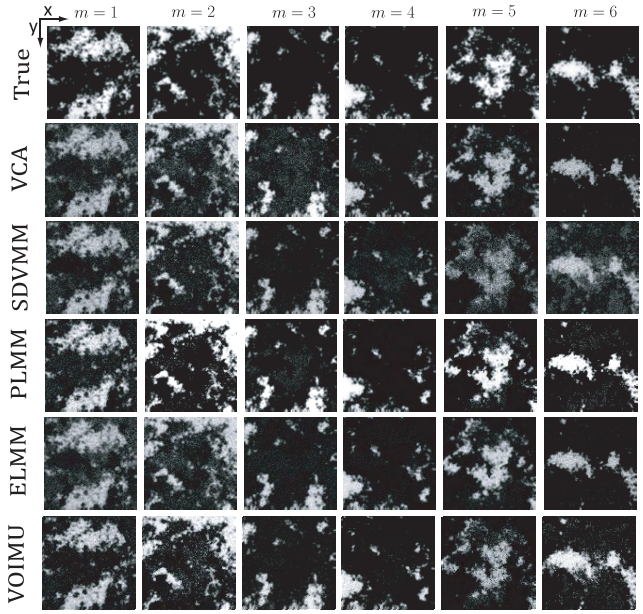

FIGURE 2. A typical realization of six estimated abundance maps by the algorithms under test for Scenario 1 (with EV without OE) together with the six true abundances (the top row), where the gray level scale is between 0 (black) and 1 (white).

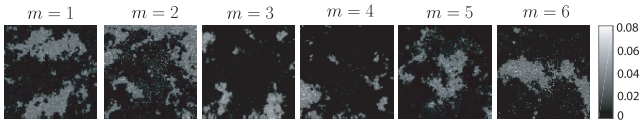

FIGURE 3. The corresponding square root of EV energy distribution (cf. (31)) associated with the VOIMU algorithm for the same realization in Figure 2.

is shown in Figure 3, which also shares the same distribution pattern as the associated abundance map (cf. Figure 2) because the EV energy of each endmember only applies to those pixels where the associated material is extracted (cf. Remark 4). On the other hand, the computational costs in terms of running time for the three HU algorithms with EV taken into account are much higher than those of VCA/FCLS and SDVMM-RASF/FCLS, because when EV is considered, the number of variables to be estimated is significantly larger.

Now let us focus on the simulation results shown in Table 1 for Scenario 2 which is a more challenging scenario (in the presence of both EV and OE). Some observations are as follows. All the performances of all the algorithms under test are worse than those shown in Scenario 1, except for VOIMU and SDVMM-RASF/FCLS. Nevertheless, the VOIMU algorithm performs much better than all the other algorithms in terms of RE and xSAM. In terms of aRMSE, SAE, and AAE, the VOIMU algorithm performs slightly better than the SDVMM-RASF/FCLS algorithm (that is also robust against outliers by design), and they significantly outperform the other algorithms. This can be further justified from the estimated abundance maps that are shown in Figure 4 for a typical realization. From AAE and Figure 4, one can

TABLE 1. Simulation results for Scenario 1 (with EV without OE for SNR = 30 dB) and Scenario 2 (with both EV and OE for SNR = 30 dB, SOR = −10 dB and $Z = 10$ outliers), where the boldface numbers denote the best performance.

Methods	Scenario 1						Scenario 2					
	Performance Measures					T	Performance Measures					T
	RE	xSAM	aRMSE	SAE	AAE	(seconds)	RE	xSAM	aRMSE	SAE	AAE	(seconds)
VCA/FCLS	1.03e-03	4.18e-02	5.18e-02	2.93	15.97	6.26	2.52e-03	6.41e-02	1.30e+00	11.9	56.06	**6.31**
SDVMM-RASF/FCLS	7.53e-04	3.75e-02	5.58e-02	2.89	17.81	**6.22**	7.53e-04	3.75e-02	5.58e-02	2.89	18.14	7.33
PLMM	1.69e-04	1.89e-02	5.12e-02	1.97	12.87	1176.64	1.83e-04	1.96e-02	1.28e+00	14.0	68.70	2250.04
ELMM	8.31e-06	4.01e-03	5.64e-02	2.90	**12.33**	188.37	1.76e-05	5.49e-03	1.27e+00	11.9	53.47	413.92
VOIMU	**9.75e-07**	**7.70e-04**	**4.02e-02**	**1.09**	13.65	343.15	**9.81e-07**	**7.71e-04**	**4.02e-02**	**1.09**	**14.02**	353.84

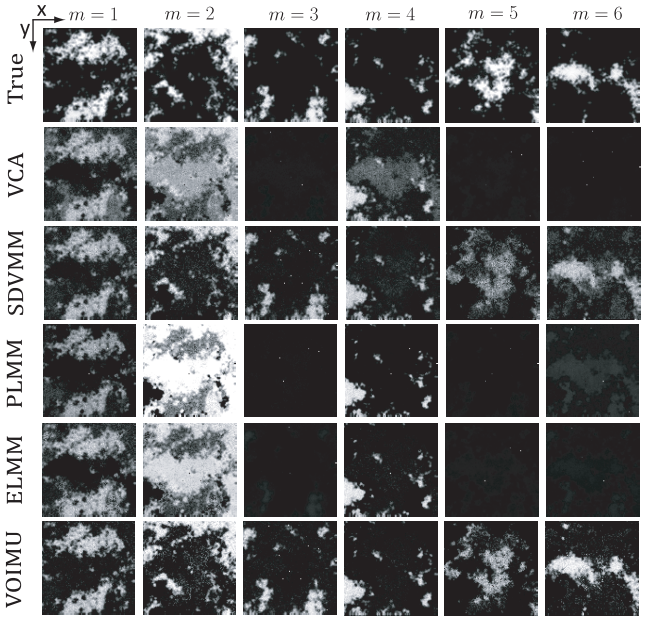

FIGURE 4. A typical realization of six estimated abundance maps by the algorithms under test for Scenario 2 (with both EV and OE) together with the six true abundances (the top row), where the gray level scale is between 0 (black) and 1 (white).

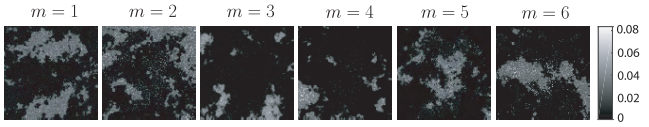

FIGURE 5. The corresponding square root of EV energy distribution (cf. (31)) associated with the VOIMU algorithm for the same realization in Figure 4.

clearly see that the VOIMU algorithm can yield more accurate abundance maps than SDVMM-RASF/FCLS, however, VCA/FCLS, PLMM and ELMM algorithms fail to yield reliable abundance maps. Surely, the distribution pattern for the square root of EV energy shown in Figure 5 and that for the estimated abundance maps shown in Figure 4 are almost the same, meaning that VOIMU algorithm is indeed outlier-insensitive. Moreover, the distribution of the normalized $1/z_n^\star$ in Scenario 1 is quite uniform, indicating no outliers in this scenario. The normalized $1/z_n^\star$ in Scenario 2 is shown in Figure 6, together with true outliers generated in Step (S4) (cf. (30)), indicating that they are highly coincident

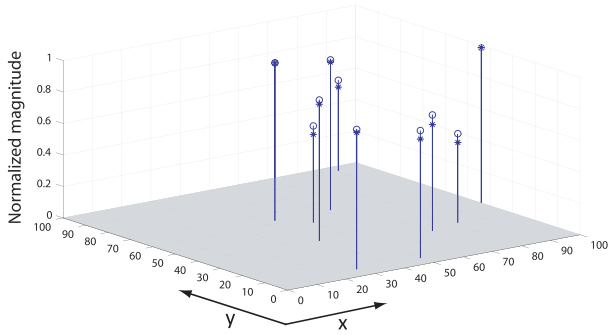

FIGURE 6. The distribution of the normalized $1/z_n^\star$ exceeding a threshold (denoted as circles) and true artificial outliers (denoted as "$*$") for the same realization in Figure 4.

with each other in outlier locations and relative magnitudes as analyzed in Remark 3. Actually, taking a close look at the abundance map for $m = 3$ associated with the VOIMU algorithm in Figure 4, all the ten outliers can be identified and they are consistent with the true outlier locations shown in Figure 6, and meanwhile the corresponding square root of EV energies in Figure 5 at the outlier locations are invisibly small (cf. Remark 3). Finally, the running times of all the algorithms under test are larger for Scenario 2 than those for Scenario 1 with different degree.

V. EXPERIMENTAL RESULTS

The proposed VOIMU algorithm is evaluated using two real datasets taken from the hyperspectral imaging data collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [44]. The experiment for each dataset is performed for two cases as follows:

- Case 1: No artificial outliers added to the original dataset;
- Case 2: 10 artificial outliers (according to (30)) added to the original dataset.

The proposed VOIMU algorithm is tested together with those algorithms for performance comparison as presented in Section IV. Again, all the parameters used for the VOIMU algorithm in the experiment are the same as used in the Subsection IV-C, and we also tried our best in finding the proper parameters for best performances of the other algorithms under test. Next, we present the experimental results.

TABLE 2. Moffett field experimental results for Case 1 (without artificial outliers) and Case 2 (with 10 artificial outliers added), where the boldface numbers denote the best performance.

Methods	Case 1			Case 2		
	Performance Measures		Running Time	Performance Measures		Running Time
	RE	xSAM	T (seconds)	RE	xSAM	T (seconds)
VCA/FCLS	1.58e-04	1.46e-01	1.96	2.39e-02	4.36e-01	**1.79**
SDVMM-RASF/FCLS	1.14e-04	9.06e-02	**1.67**	1.14e-04	9.06e-02	**1.79**
PLMM	4.36e-05	4.53e-02	6.40	7.10e-05	5.80e-02	30.49
ELMM	1.53e-05	3.33e-02	52.36	1.02e-04	7.02e-02	39.01
VOIMU	**2.80e-06**	**2.48e-03**	68.67	**2.80e-06**	**2.49e-03**	66.67

FIGURE 7. Sub-image of Moffett field.

A. SUBSCENE OF MOFFETT FIELD

A sub-image of size 50 lines by 50 columns taken from the hyperspectral image of Moffett Field (cf. Fig. 7), consisting of 224 spectral bands. With water absorption bands removed, the left 189 exploitable spectral bands are used in the experiment. This sub-image has been extensively studied in HU literature [16], in which the number of endmembers is known as $N = 3$, including vegetation, water, and soil. The parameters used for PLMM are $(\alpha, \beta, \gamma) = (0.05, 0, 1)$ and those used for ELMM are $(\lambda_S, \lambda_A, \lambda_\psi) = (0.4, 5 \times 10^{-3}, 10^{-3})$.

The experimental results in terms of RE, xSAM and running time for Case 1 and Case 2 are shown in Table 2. The estimated three abundance maps are shown in Figure 8 and Figure 9 for Case 1 and Case 2, respectively. From the results for Case 1, one can observe that the VOIMU algorithm significantly outperforms all the other algorithms, ELMM performs slightly better than PLMM, and PLMM performs much better than VCA/FCLS and SDVMM-RASF/FCLS, and that the estimated abundance maps are also quite similar to each other (cf. Figure 8). These observations are also consistent with those from the simulation results for Scenario 1 in the previous section (simulation results).

From the results for Case 2 in Table 2, one can observe that the VOIMU algorithm still outperforms all the other algorithms. It is noticeable by comparing Figure 8 and Figure 9 that the abundance maps associated with VOIMU and SDVMM-RASF/FCLS algorithms are much more reliable than those associated with the other algorithms, and OE has larger impact on the HU performance than EV. These observations are also consistent with those from the simulation results for Scenario 2 in the previous simulation section. Moreover, the square root of EV energy

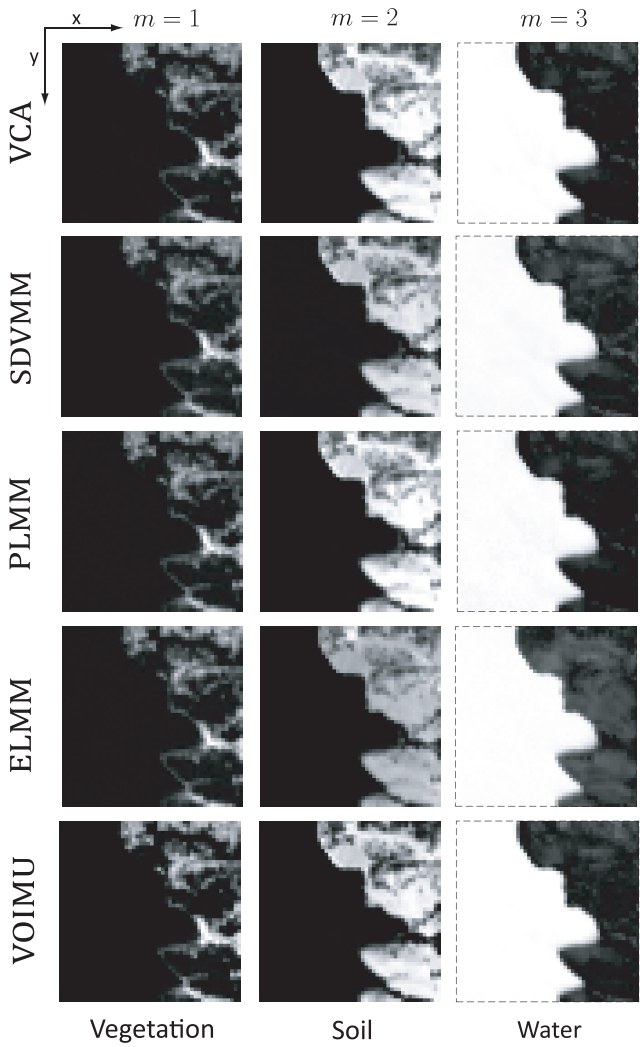

FIGURE 8. Three estimated abundance maps of the Moffett field sub-image by the algorithms under test for Case 1 (without artificial outliers), where the gray level scale is between 0 (black) and 1 (white).

distributions of VOIMU for Case 1 and Case 2 are shown in Figures 10(a) and 10(b), respectively. The three estimated endmember signatures $\widehat{\mathbf{A}}_n$ and the estimated reference signatures $\bar{\mathbf{A}}$ by VOIMU are shown in Figure 11(a) for Case 1, and Figure 11(b) for Case 2. Note that the results shown in Figures 10 are similar to each other for the two cases and so are those shown in Figure 11, demonstrating that the VOIMU

TABLE 3. Cuprite Desert Varnish sub-image experimental results for Case 1 (without artificial outliers) and Case 2 (with 10 artificial outliers added), where the boldface numbers denote the best performance.

Methods	Case 1			Case 2		
	Performance Measures		Running Time	Performance Measures		Running Time
	RE	xSAM	T (seconds)	RE	xSAM	T (seconds)
VCA/FCLS	1.54e-03	5.45e-02	3.08	1.07e-01	1.43e-01	**3.11**
SDVMM-RASF/FCLS	6.76e-05	1.42e-02	**2.96**	6.76e-05	1.42e-02	3.57
PLMM	6.72e-05	1.27e-02	82.37	5.57e-05	1.40e-02	741.98
ELMM	1.67e-05	6.25e-03	34.80	3.40e-05	1.04e-02	62.01
VOIMU	**3.79e-08**	**2.82e-04**	129.03	**3.74e-08**	**2.82e-04**	124.88

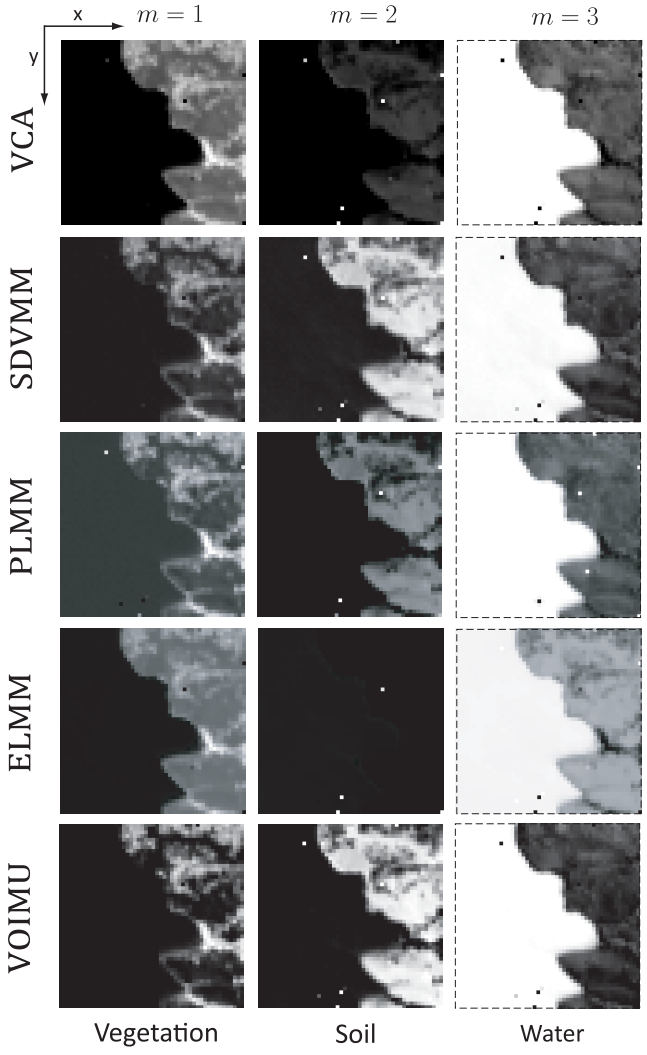

FIGURE 9. Three estimated abundance maps of the Moffett field sub-image by the algorithms under test for Case 2 (with 10 artificial outliers), where the gray level scale is between 0 (black) and 1 (white).

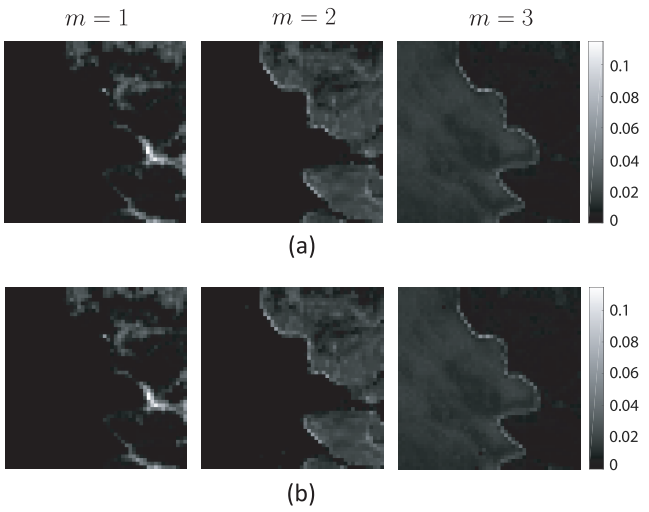

FIGURE 10. The corresponding square root of EV energy distribution of the Moffett field sub-image associated with the VOIMU algorithm for (a) Case 1 and (b) Case 2.

algorithm is robust against outliers. The normalized $1/z_n^\star$ exceeding a threshold (denoted by circles) for Case 1 is shown in Figure 12(a), and that for Case 2 in Figure 12(b) as well as the true artificial outliers (denoted as "$*$"). It can be observed that some moderate outliers appear in Figure 12(a) and they can also be seen in Figure 12(b) in addition to high coincidence between the distribution of $1/z_n^\star$ (excluding the moderate outliers) and the true artificial outliers. Again, these observations are also consistent with those from the simulations results for Scenario 2. Finally, the running times of VOIMU are similar for the two cases (cf. Table 2).

B. SUB-IMAGE OF CUPRITE: DESERT VARNISH

A sub-image of size 50 lines by 90 columns taken from the hyperspectral image of well-known AVIRIS Cuprite (cf. Fig. 13), with water absorption bands removed (1-3, 104-113, 148-168, and 221-224), the left 187 exploitable spectral bands are used in the experiment. This sub-image also extensively studied in HU literature, is known to have $N = 3$ endmembers, including Montmorillonite, Desert Varnish, and Alunite [45], [46]. The parameters used for PLMM and ELMM are $(\alpha, \beta, \gamma) = (0.014, 404, 1)$ and $(\lambda_S, \lambda_A, \lambda_\psi) = (0.4, 5\times10^{-3}, 5\times10^{-3})$, respectively.

The experimental results in terms of RE, xSAM and running time for Case 1 and Case 2 are shown in Table 3. The estimated three abundance maps are shown in Figure 14 and Figure 15 for Case 1 and Case 2, respectively. From Table 3, one can observe that for Case 1, the VOIMU algorithm significantly outperforms all the other algorithms, whose performances are basically comparable and thus their

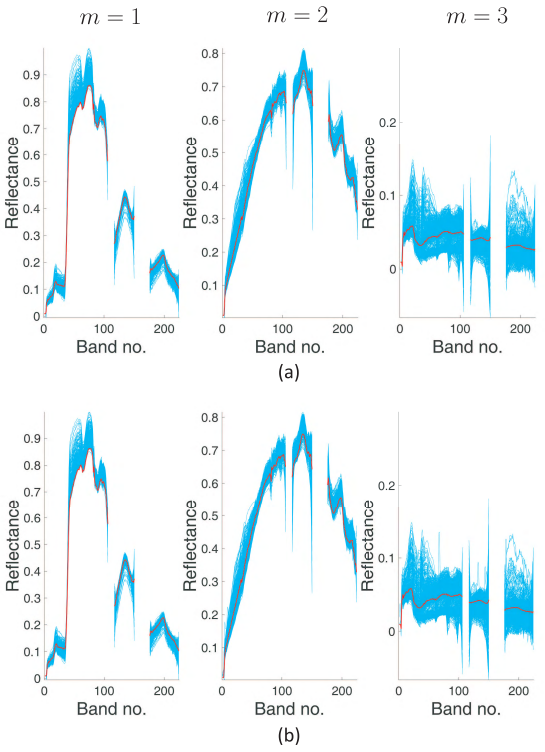

FIGURE 11. Three estimated endmember signatures $\widehat{\mathbf{A}}_n$ (blue line) and the estimated reference signatures $\bar{\mathbf{A}}$ (red line) of the Moffett field sub-image by the VOIMU algorithm for (a) Case 1, and (b) Case 2.

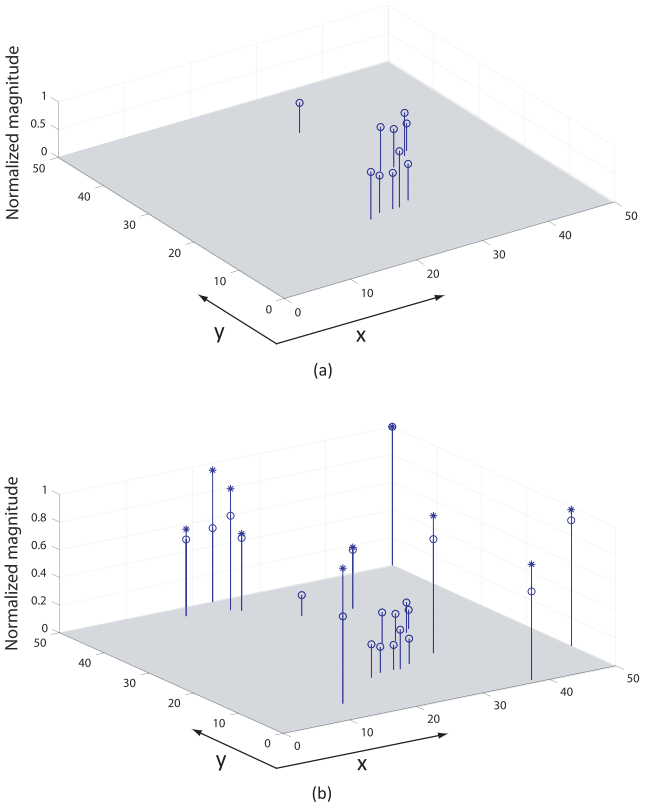

FIGURE 12. The distribution of the normalized $1/z_n^\star$ exceeding a threshold (denoted as circles) of the Moffett field sub-image for (a) Case 1 and (b) Case 2, where true artificial outliers are denoted as "$*$".

estimated abundance maps are somewhat similar as well (cf. Figure 14).

From Case 2 in Table 3, one can observe that the VOIMU algorithm significantly outperforms the other algorithms. By comparing Figure 14 and Figure 15, we can see that the abundance maps for VOIMU and SDVMM-RASF/FCLS are much more reliable than those obtained by the other algorithms under test for this case, and again, OE has larger impact on the HU performance than EV. These observations are also consistent with those from the simulation results for Scenario 2 in Section IV. Moreover, the square root of EV energy distributions of VOIMU for Case 1 and Case 2 are shown in Figures 16(a) and 16(b), respectively. The three estimated endmember signatures $\widehat{\mathbf{A}}_n$ and the estimated reference signatures $\bar{\mathbf{A}}$ by VOIMU are shown in Figure 17(a) for Case 1, and Figure 17(b) for Case 2. Note that the results shown in Figure 16 are similar to each other for the two cases and so are those shown in Figure 17, demonstrating that VOIMU is robust against outliers. The distribution of $1/z_n^\star$ in Case 1 is quite uniform, thus indicating no non-negligible outliers in the original Cuprite Desert Varnish dataset. Moreover, the normalized $1/z_n^\star$ exceeding a threshold (denoted by circles) for Case 2 is shown in Figure 18 as well as the true artificial outliers (denoted as "$*$"), which can also be identified from the abundance maps associated with the VOIMU algorithm for $m = 2$ and $m = 3$ in Figure 15. Again, these observations are also consistent with those from the simulations results

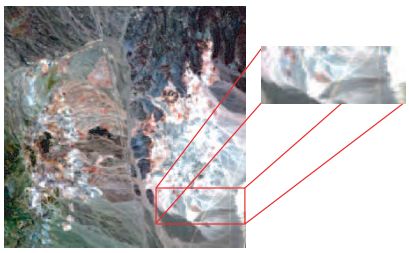

FIGURE 13. Sub-image of Cuprite: Desert Varnish.

for Scenario 2 in Section IV. Finally, the running times of VOIMU are also similar for the two cases (cf. Table 3).

The presented experimental study only for $N = 3$ above is for ease of demonstrating the effectiveness of the proposed algorithm (e.g., outlier detectability) in a visually concise and clear manner. Some more experimental results using AVIRIS data for $N = 9$ are provided in a separate but self-contained technical report in a GitHub link[1], to further support the practical applicability of the proposed algorithm.

[1] https://github.com/roy50408/VOIMU

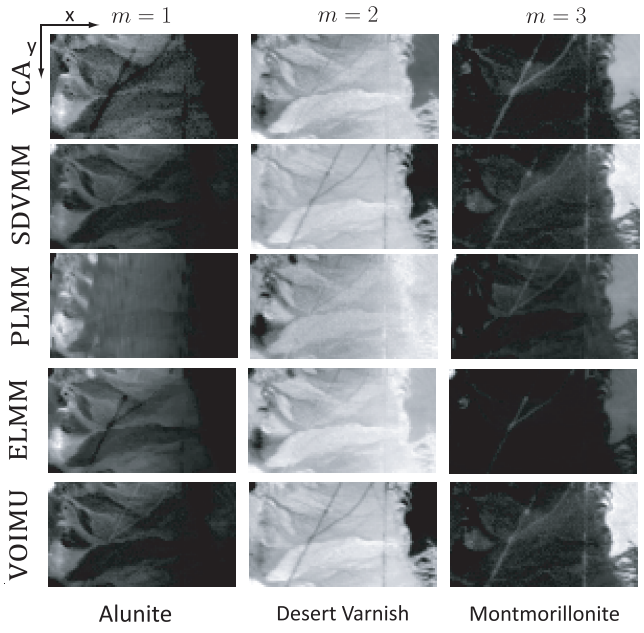

FIGURE 14. Three estimated abundance maps of the Cuprite Desert Varnish sub-image by the algorithms under test for Case 1 (without artificial outliers), where the gray level scale is between 0 (black) and 1 (white).

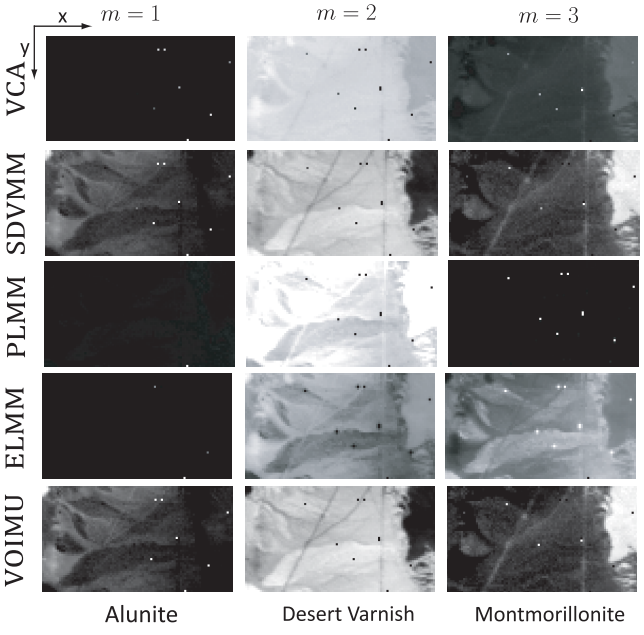

FIGURE 15. Three estimated abundance maps of the Cuprite Desert Varnish sub-image by the algorithms under test for Case 2 (with 10 artificial outliers), where the gray level scale is between 0 (black) and 1 (white).

VI. CONCLUSION

Motivated by the robustness property of p-quasi norm to outliers (where $0 < p < 1$), we have presented a PLMM based HU algorithm, i.e., the VOIMU algorithm, that is robust against EV and OE in the meantime, by solving a nonconvex minimization problem as given by (4), where an

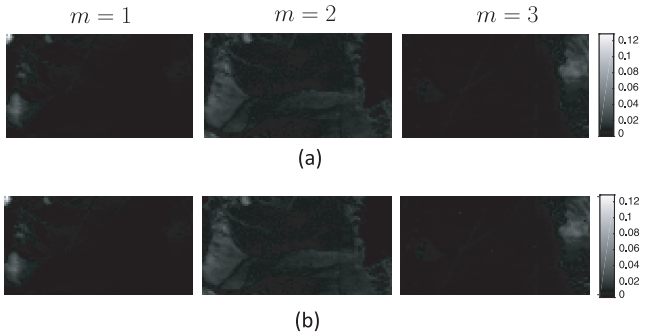

FIGURE 16. The corresponding square root of EV energy distribution of the Cuprite Desert Varnish sub-image associated with the VOIMU algorithm for (a) Case 1 and (b) Case 2.

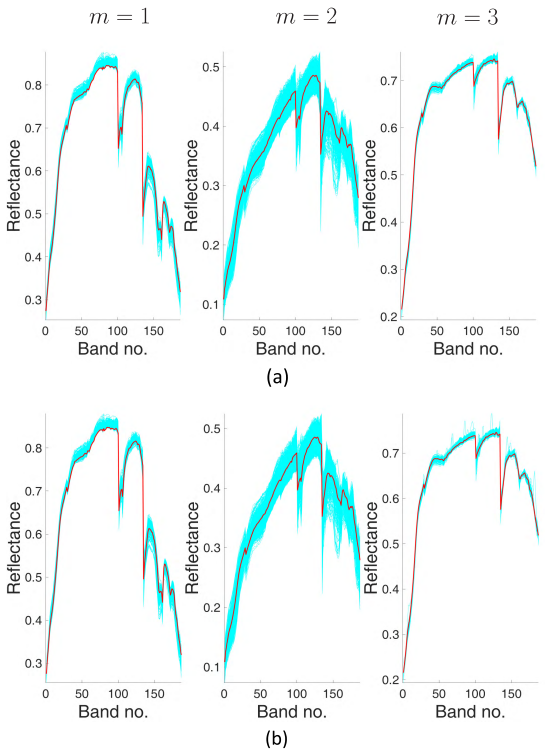

FIGURE 17. Three estimated endmember signatures $\widehat{\mathbf{A}}_n$ (blue line) and the estimated reference signatures $\bar{\mathbf{A}}$ (red line) of the Cuprite Desert Varnish sub-image by the VOIMU algorithm for (a) Case 1, and (b) Case 2.

EV energy based regularizer and a simplex volume based regularizer are used in order to get the solution with desired characteristics. Then problem (4) is reformulated into a multi-convex problem as given by (7), with an auxiliary variable $\mathbf{z} \in \mathbb{R}_+^L$ used for optimal weights of the data-fitting errors. The proposed VOIMU algorithm is designed by applying the BCD method to solve (7), and implemented by Algorithm 1, which can yield a stationary point of (7) with convergence guarantee. A remarkable property of Algorithm 1 is that the yielded solution for $\mathbf{z}$ exhibits potential outlier locations (cf. Remark 3). Some simulation results and experimental results using AVIRIS data have been provided to demonstrate the effectiveness of the proposed VOIMU algorithm, and its

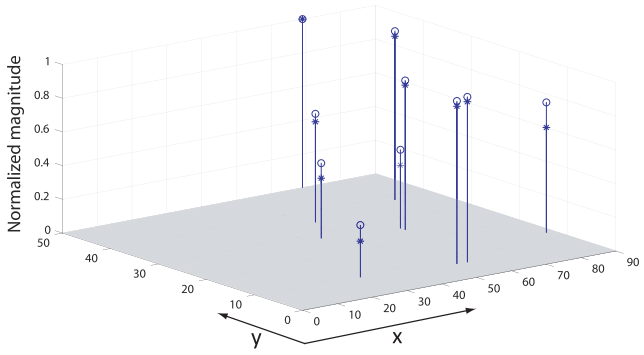

FIGURE 18. The distribution of the normalized $1/z_n^\star$ exceeding a threshold (denoted as circles) of the Cuprite Desert Varnish sub-image for Case 2, where true artificial outliers are denoted as "$*$".

superior performance over some benchmark physics-model based algorithms that are either robust against EV or robust against OE. However, no performance comparison with any existing HU algorithms that can handle both EV and OE at the same time is presented because of no such algorithms reported in the literature as far as we know. How to upgrade the computational efficiency of the VOIMU algorithm is left as a future study.

APPENDIX A
PROOF OF PROPOSITION 1

Let $\boldsymbol{x}_1 = \mathbf{z}$, $\boldsymbol{x}_2 = [\mathbf{vec}(\mathbf{A}_1)^T, \ldots, \mathbf{vec}(\mathbf{A}_L)^T]^T$, $\boldsymbol{x}_3 = \mathbf{vec}(\bar{\mathbf{A}})$, $\boldsymbol{x}_4 = \mathbf{vec}(\mathbf{S})$, and $\boldsymbol{x} = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_4^T]^T$. Let $f(\boldsymbol{x})$ be the objective function of problem (7) and $\mathcal{X}$ be the associated feasible set. Note that f is differentiable [47] and nonconvex though multi-convex, and $\mathcal{X}$ is a closed convex set. Let

$$\bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}}) = f(\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_{i-1}, \boldsymbol{x}_i, \bar{\boldsymbol{x}}_{i+1}, \ldots, \bar{\boldsymbol{x}}_4), \tag{32}$$

and $\mathcal{X}_i$, $i = 1, \ldots, 4$, be the objective functions and the associated feasible sets of convex subproblems (9), (10), (11), and (12), respectively, where $\bar{\boldsymbol{x}} \in \mathcal{X}$ denotes a given reference point. Let

$$\bar{\boldsymbol{\xi}}_i = \arg\min_{\boldsymbol{x}_i \in \mathcal{X}_i} \bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}}), \tag{33}$$

which is unique because $\bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}})$ is strictly convex [48], [49] and $\mathcal{X}_i$ is closed convex for all $i = 1, \ldots, 4$.

Algorithm 1 is an instance of the iterative block successive upper bound minimization (BSUM) method [48], [49], that at the r-th iteration, $\bar{\boldsymbol{x}}_i$ is updated by $\bar{\boldsymbol{x}}_i = \bar{\boldsymbol{\xi}}_i$ in a round-robin fashion, where $i = ((r-1) \bmod 4) + 1$. It can be seen that Algorithm 1 also satisfies all the convergence conditions required for obtaining a stationary-point solution by BSUM when f and $\bar{f}_i$ are differentiable as follows [49, Remark 4.10]:

$\bar{f}_i(\bar{\boldsymbol{x}}_i \mid \bar{\boldsymbol{x}}) = f(\bar{\boldsymbol{x}})$, (since (32))

$\bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}}) \geq f(\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_{i-1}, \boldsymbol{x}_i, \bar{\boldsymbol{x}}_{i+1}, \ldots, \bar{\boldsymbol{x}}_4)$, (since (32))

$\bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}})$ is continuous in $(\boldsymbol{x}_i, \bar{\boldsymbol{x}})$, (as f is differentiable)

problem (33) has a unique solution, (as mentioned above)

as well as the premise that $\bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}})$ is quasiconvex in $\boldsymbol{x}_i$, $\forall i = 1, \ldots, 4$ (since $\bar{f}_i(\boldsymbol{x}_i \mid \bar{\boldsymbol{x}})$ is strictly convex), and $f(\boldsymbol{x})$ is regular at every point $\boldsymbol{x} \in \mathcal{X}$ (since f is differentiable). Therefore, the solution obtained by Algorithm 1 is guaranteed to be a stationary point of problem (7). ■

APPENDIX B
PROOF OF LEMMA 1

First of all, let us prove that $f_\varepsilon(v, z)$ given by (5) is convex in z by the second-order condition. It can be easily derived that

$$\frac{\partial f_\varepsilon(v,z)}{\partial z} = 2v^2 z + \alpha \cdot \frac{2p}{p-2} \cdot z^{\frac{p+2}{p-2}} + 2\varepsilon z,$$
$$\frac{\partial^2 f_\varepsilon(v,z)}{\partial z^2} = 2v^2 + \alpha \cdot \frac{2p(p+2)}{(p-2)^2} \cdot z^{\frac{4}{p-2}} + 2\varepsilon.$$

Because of $\partial^2 f_\varepsilon(v,z)/\partial z^2 > 0$ for all $z \geq 0$, $f_\varepsilon(v,z)$ is strictly convex in z. Next, let $\tau = (2-p)(v^2+\varepsilon)/\alpha p$, to find the optimal z, by setting $\partial f_\varepsilon(v,z)/\partial z = 0$, and then we come up with the optimal $z^\star > 0$ given by (14) and the associated optimal value is given by

$$\begin{aligned} f_\varepsilon(v, z^\star) &= g_\varepsilon(z^\star)v^2 + h_\varepsilon(z^\star) \\ &= \tau^{(p-2)/2}(v^2+\varepsilon) + \alpha\tau^{p/2} \\ &= \Big(\frac{2\alpha}{2-p}\Big)\Big(\frac{2-p}{\alpha p}\Big)^{p/2}(v^2+\varepsilon)^{p/2} \\ &= (v^2+\varepsilon)^{p/2}, \end{aligned}$$

where lengthy mathematical manipulations were performed in the derivations of the third and fourth equalities. Thus we have completed the proof. ■

APPENDIX C
PROOF OF PROPOSITION 2

By (18c), we have

$$\boldsymbol{\Phi}_j^T \boldsymbol{\Phi}_j = \mathbf{DIAG}\{(z_n^{k+1})^2 \mathbf{s}_n^k (\mathbf{s}_n^k)^T\}_{n=(j-1)L_s+1}^{jL_s} \otimes \mathbf{I}_M. \tag{34}$$

Then substituting (34) into the first term in (22) followed by some matrix manipulations, we have

$$(\boldsymbol{\Phi}_j^T \boldsymbol{\Phi}_j + (\lambda_1 + \eta)\mathbf{I}_{MNL_s})^{-1} = \boldsymbol{\Psi} \otimes \mathbf{I}_M, \tag{35}$$

where $\boldsymbol{\Psi}$ is defined in (24a). Next, by substituting (35) into (22), we come up with

$$\begin{aligned} \mathbf{c}_j^{i+1} &= (\boldsymbol{\Psi} \otimes \mathbf{I}_M) \times (\boldsymbol{\Phi}_j^T \mathbf{q}_j + \boldsymbol{v}), \\ &= \mathbf{vec}(\mathbf{devec}(\boldsymbol{\Phi}_j^T \mathbf{q}_j + \boldsymbol{v}, M, NL_s) \times \boldsymbol{\Psi}) \\ &= \mathbf{vec}\big((\mathbf{devec}(\boldsymbol{\Phi}_j^T \mathbf{q}_j, M, NL_s) + \boldsymbol{\Gamma}) \times \boldsymbol{\Psi}\big) \\ &= \mathbf{vec}\Big(\big(\mathbf{devec}(\mathbf{q}_j, M, L_s)\boldsymbol{\Lambda} + \boldsymbol{\Gamma}\big)\boldsymbol{\Psi}\Big) \end{aligned} \tag{36}$$

which is exactly (23), and $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are defined in (24b) and (24c), respectively. In the derivation of (36), the second equality is due to the Kronecker product property of $(\mathbf{U}^T \otimes \mathbf{V})\mathbf{vec}(\mathbf{Y}) = \mathbf{vec}(\mathbf{VYU})$ and the fact that $\boldsymbol{\Psi}$ is symmetric; the last equality is obtained by applying the same

Kronecker product property as used in the third equality and the Kronecker product structure of $\mathbf{\Phi}_j$ (cf. (18c)) to $\mathbf{\Phi}_j^T \mathbf{q}_j$. Hence, the proof of Proposition 2 is completed. ■

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, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[3] W.-K. Ma, J. M. Bioucas-Dias, J. Chanussot, and P. Gader, "Signal and image processing in hyperspectral remote sensing [from the guest editors]," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 22–23, Jan. 2014.

[4] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[5] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.

[6] A. Plaza, P. Martínez, R. Pérez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.

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

[8] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[9] C.-H. Lin, R. Wu, W.-K. Ma, C.-Y. Chi, and Y. Wang, "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization," *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 1651–1679, Jun. 2018.

[10] C.-H. Lin and J. M. B. Dias, "Provably and roubst blind source seperation of ill-conditioned hyperspectral mixtures," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2018, pp. 333–337.

[11] C.-H. Lin and J. M. B. Dias, "New theory for unmixing ill-conditioned hyperspectral mixtures," in *Proc. IEEE 10th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2018, pp. 430–434.

[12] A. Zare and K. C. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[13] B. Zhang, L. Zhuang, L. Gao, W. Luo, Q. Ran, and Q. Du, "PSO-EM: A hyperspectral unmixing algorithm based on normal compositional model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7782–7792, Dec. 2014.

[14] X. Du, A. Zare, P. Gader, and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1994–2003, Jun. 2014.

[15] Y. Zhou, A. Rangarajan, and P. D. Gader, "A Gaussian mixture model representation of endmember variability in hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2242–2256, May 2018.

[16] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.

[17] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.

[18] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[19] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.

[20] T.-H. Chan, A. Ambikapathi, W.-K. Ma, and C.-Y. Chi, "Robust affine set fitting and fast simplex volume max-min for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3982–3997, Jul. 2013.

[21] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012.

[22] B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[23] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[24] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[25] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6254–6268, Dec. 2016.

[26] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2678–2689, Aug. 2005.

[27] H. Ekblom, "L_p-methods for robust regression," *BIT Numer. Math.*, vol. 14, no. 1, pp. 22–32, 1974.

[28] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, Jun. 2001.

[29] C.-H. Lin, C.-Y. Chi, L. Chen, D. J. Miller, and Y. Wang, "Detection of sources in non-negative blind source separation by minimum description length criterion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4022–4037, Sep. 2018.

[30] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[31] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1946–1961, Apr. 2016.

[32] 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.

[33] C.-H. Lin, F. Ma, C.-Y. Chi, and C.-H. Hsieh, "A convex optimization-based coupled nonnegative matrix factorization algorithm for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1652–1667, Mar. 2018.

[34] C.-H. Lin, A. Ambikapathi, W.-C. Li, and C.-Y. Chi, "On the endmember identifiability of Craig's criterion for hyperspectral unmixing: A statistical analysis for three-source case," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 2139–2143.

[35] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, Oct. 2015.

[36] C.-H. Lin and J. M. Bioucas-Dias, "Linear spectral unmixing via matrix factorization: Identifiability criteria for sparse abundances," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Valencia, Spain, Jul. 2018, pp. 6155–6158.

[37] 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.

[38] M. Grant, S. Boyd, and Y. Ye. (Apr. 2011). *CVX: MATLAB Software for Disciplined Convex Programming, Version 1.21*. [Online]. Available: http://cvxr.com/cvx/

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

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

[41] R. N. Clark *et al.* (2007). *USGS Digital Spectral Library Splib06a: U.S. Geological Survey, Digital Data Series 231*. [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

[42] 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.
[43] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, Oct. 2016.
[44] *AVIRIS Free Standard Data Products*. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html
[45] G. Swayze, R. N. Clark, F. Kruse, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, vol. 1, 1992, pp. 47–49.
[46] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, Mar. 2012.
[47] R. L. Wheeden and A. Zygmund, *Measure and Integral: An Introduction to Real Analysis*. New York, NY, USA: Marcel Dekker, 1977.
[48] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
[49] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. Boca Raton, FL, USA: CRC Press, 2017.

YAO-RONG SYU received the B.S. degree from the Communications Engineering Department, National Central University, Taiwan, in 2015. He is currently pursuing the master's degree with the Department of Electrical Engineering, Institute of Communications Engineering, National Tsing Hua University. His research interests include hyperspectral unmixing, convex geometry and optimization, fast algorithm design, data fusion, and blind source separation.

CHIA-HSIANG LIN received the B.S. degree in electrical engineering and the Ph.D. degree in communications engineering from National Tsing Hua University (NTHU), Taiwan, in 2010 and 2016, respectively. From 2015 to 2016, he was a Visiting Student with Virginia Tech, VA, USA.

He held research positions with NTHU, from 2016 to 2017, The Chinese University of Hong Kong, Hong Kong, in 2014 and 2017, and the University of Lisbon, Portugal, from 2017 to 2018. He was an Assistant Professor with the Center for Space and Remote Sensing Research, National Central University, Taiwan, in 2018. He is currently an Assistant Professor with the Department of Electrical Engineering, and also with the Institute of Computer and Communication Engineering, National Cheng Kung University, Taiwan, both since 2019. His research interests include network science, game theory, convex geometry and optimization, blind source separation, and imaging science.

Dr. Lin received the Ministry of Science and Technology (MOST) Young Scholar Fellowship, Taiwan, together with the Einstein Grant Award, from 2018 to 2023. In 2016, he was a recipient of the Outstanding Doctoral Dissertation Award from the Chinese Image Processing and Pattern Recognition Society and the Best Doctoral Dissertation Award from the IEEE Geoscience and Remote Sensing Society.

CHONG-YUNG CHI received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1983. From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA, USA. He has been a Professor with the Department of Electrical Engineering, since 1989, and has also been with the Institute of Communications Engineering (ICE), since 1999. He was the Chairman of ICE, from 2002 to 2005, and National Tsing Hua University, Hsinchu, Taiwan. He has published more than 230 technical papers (with citations more than 4500 times by Google-Scholar), including more than 85 journal papers (mostly in the IEEE TRANSACTIONS ON SIGNAL PROCESSING), more than 140 peer-reviewed conference papers, four book chapters, and two books, including a new textbook, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications* (CRC Press, 2017) [which has been popularly used in an invited intensive (2 or 3 weeks) short course 17 times in nine major universities in China, since 2010, before its publication]. Recently, he received the 2018 IEEE Signal Processing Society Best Paper Award. His current research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, and biomedical and hyperspectral image analysis.

He is a Senior Member of the IEEE. He was a member of the Signal Processing Theory and Methods Technical Committee, from 2005 to 2010, the Signal Processing for Communications and Networking Technical Committee, from 2011 to 2016, and the Sensor Array and Multichannel Technical Committee, from 2013 to 2018, the IEEE Signal Processing Society. He has been a Technical Program Committee Member for many IEEE sponsored and co-sponsored workshops, symposiums, and conferences on signal processing and wireless communications, including the 2001 IEEE Workshop on Signal Processing Advances in Wireless Communications as a Co-Organizer and a General Co-Chairman, a Co-Chair of Signal Processing for Communications (SPC) Symposium, ChinaCOM 2008, and a Lead Co-Chair of SPC Symposium, ChinaCOM 2009. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, from 2001 to 2006 and from 2012 to 2015, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, from 2006 to 2007, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, from 2008 to 2009, and the IEEE SIGNAL PROCESSING LETTERS, from 2006 to 2010. He was a member of the Editorial Board of Elsevier *Signal Processing* (6/2005–5/2008), an Editor and a Guest Editor of *EURASIP Journal on Applied Signal Processing*, from 2003 to 2005, and in 2006, respectively.