Robust Affine Set Fitting and Fast Simplex Volume Max-Min for Hyperspectral Endmember Extraction

Tsung-Han Chan, *Member, IEEE*, ArulMurugan Ambikapathi, *Member, IEEE*, Wing-Kin Ma, *Senior Member, IEEE*, and Chong-Yung Chi, *Senior Member, IEEE*

Abstract—Hyperspectral endmember extraction is to estimate endmember signatures (or material spectra) from the hyperspectral data of an area for analyzing the materials and their composition therein. The presence of noise and outliers in the data poses a serious problem in endmember extraction. In this paper, we handle the noise- and outlier-contaminated data by a two-step approach. We first propose a robust-affine-set-fitting algorithm for joint dimension reduction and outlier removal. The idea is to find a contamination-free data-representative affine set from the corrupted data, while keeping the effects of outliers minimum, in the least squares error sense. Then, we devise two computationally efficient algorithms for extracting endmembers from the outlier-removed data. The two algorithms are established from a simplex volume max-min formulation which is recently proposed to cope with noisy scenarios. A robust algorithm, called worst case alternating volume maximization (WAVMAX), has been previously developed for the simplex volume max-min formulation but is computationally expensive to use. The two new algorithms employ a different kind of decoupled max-min partial optimizations, wherein the design emphasis is on low-complexity implementations. Some computer simulations and real data experiments demonstrate the efficacy, the computational efficiency, and the applicability of the proposed algorithms, in comparison with the WAVMAX algorithm and some benchmark endmember extraction algorithms.

Index Terms—Alternating optimization, fast endmember extraction, hyperspectral images, robust dimension reduction, simplex volume max-min, successive optimization.

I. INTRODUCTION

HYPERSPECTRAL remote sensing exploits the fact that all substances uniquely reflect, absorb, and emit electromagnetic (EM) energy, at specific wavelength, in distinctive patterns depending on their molecular composition. The hyperspectral sensor collects data in hundreds of narrow con-tiguous spectral bands, thereby providing a powerful means to discriminate disparate materials based on their unique spectral signatures, or *endmember signatures* (simply *endmembers*). However, depending on the spatial resolution of the hyperspectral sensor, surface patches corresponding to individual pixels may still contain more than one material [1], [2]. Hence, the decomposition of the received spectra into a set of endmembers and their corresponding mixing proportions, or *abundances*, would facilitate the characterization of different materials over the scanned area of interest [1], [2]. In the past several years, endmember extraction using hyperspectral images has been widely investigated and has proven to be valuable in many applications, including but not limited to geology, hydrology, urban planning, geography, cadastral mapping, cartography, and military [3]–[5].

The presence of noise and outliers in the hyperspectral data is inevitable in practice and may seriously affect the analysis of hyperspectral data. The noise is generated because of the random nature of the photon arrival/detection process, the sensor electronics, and quantization [6]. Raw data calibration routines could alter the statistics of the noise, so the noise features may not explicitly depend on the wavelength of the hyperspectral sensor [6], [7]. The other uncertainty is outliers. In general, the outliers are thought of as the pixels that deviate markedly from the rest of the data. Two definitions of the outlier pixels have been presented [8]–[11] in the open literature. The first refers to the pixels that provide constant or error readout, also called "dead" or "bad" pixels. Possible causes of such outlier pixels include detector failure, errors during data transfer, and improper data correction [8], [9]. The second refers to the pixels that have different spectral signatures from the background representatives. These pixels are also commonly called targets or objects in the domain of hyperspectral anomaly detection [10], [11].

Given the fact that the noise and outliers are the major source of errors in hyperspectral endmember extraction, the design of endmember extraction algorithms (EEAs) should take both the noise and outliers into account. Existing efforts that account for the noise and/or outlier effects include joint Bayesian algorithm [12], simplex identification by split augmented Lagrangian [13], robust minimum volume enclosing simplex algorithm [14], and others [15], [16], and they are carried out by different techniques. Specifically, a Bayesian estimation framework explicitly accounting for the presence of noise is employed in [12], soft constraints are utilized to mitigate outlier and/or noise pixel effects in [13], chance constraints are applied to the original minimum volume enclosing simplex constraints [17] to account for noise effects in [14], Filippi and Archibald [15] utilize a support vector machine-based approach for robustly

extracting the simplex topology, and Duran and Petrou [16] propose a robust unconstrained linear unmixing (RULU) algorithm to extract endmembers in the presence of outliers. Simply speaking, the RULU algorithm uses a clustering method for background endmember estimation and unconstrained abundance estimation, thresholding for outlier pixel removal, and data indexing with extreme abundances for robust endmember estimation. Although RULU provides robustness to outliers, the noise uncertainty still remains. Moreover, a noise-robust spatial preprocessing module has been proposed in [18], and it can be easily coupled with any nonrobust EEAs. Other than the aforementioned methods, we have also reported a robust generalization of the maximum simplex volume criterion proposed by Winter [19], [20] to account for noise effects [21] and formulated the robust Winter criterion as a simplex volume max-min problem. Although an algorithm, called worst case alternating volume maximization (WAVMAX), has been proposed to handle the robust Winter criterion in a disciplined manner, which uses a combination of alternating optimization and the subgradient method [21], it is quite computationally expensive to use for massive amounts of high-dimensional data.

In this paper, we develop robust hyperspectral EEAs for coping with outliers and noise. We first propose a robust version of the affine set fitting (ASF) [17], [22] for joint dimension reduction and outlier removal. The idea is to find a corruption-free data-representative affine set from the given hyperspectral data, while keeping the outlier effects minimum, in the least squares error sense. The proposed algorithm, called robust ASF (RASF), implements the idea by using alternating optimization. With the outlier pixels being removed by the RASF algorithm, the uncertainty that remains in the data is noise. We then propose two computationally efficient algorithms to implement the robust Winter criterion, the simplex volume max-min formulation [21], by a partial max-min optimization approach that provides closed-form solutions for either of maximization and minimization. The proposed EEAs, named alternating decoupled volume max-min (ADVMM) and successive decoupled volume max-min (SDVMM), approximate the simplex volume max-min problem by a set of decoupled max-min problems in alternating manner and successive manner, respectively. Some justifications of the decoupled max-min heuristic that motivates the development of the proposed ADVMM and SDVMM algorithms are also discussed. Simulations and experimental results will be provided to demonstrate the efficacy, the computational efficiency, and the real applicability of the proposed methods.

We should emphasize that, in this work, the noise is assumed to be zero mean isotropically distributed with identical variance over all the hyperspectral bands and assumed to be spatially homogeneous. In practical scenarios where the noise is non-isotropic, the noise prewhitening technique can be applied to the data with the noise covariance matrix estimated by the multiple regression method [23]. The consideration of the signal-dependent noise is beyond the scope of this paper. Interested readers can refer to [24] for further details. We will only focus on the outliers that are "dead" or "bad" pixels providing constant or error readout, instead of the target/object-type outliers. From the perspective of endmember extraction, the target spectra and the background spectra can both be seen as unknown endmembers, thereby being able to be readily

estimated by any EEAs [25]. Once both target and background spectra are extracted, differentiation of the target spectra from the background spectra is a separate problem [16], [26].

The outline of this paper is as follows. Section II describes the hyperspectral endmember extraction problem in the presence of noise and outliers. Section III presents the proposed RASF for dimension reduction and outlier removal. Section IV briefly reviews Winter's endmember extraction criterion and its robust generalization. Section V presents the two proposed fast algorithms for the simplex volume max-min formulation. Sections VI and VII show the results of computer simulations and real hyperspectral data experiments, respectively. Finally, some conclusions are drawn in Section VIII.

Notation: $\mathbb{R}^N$ and $\mathbb{R}^{M \times N}$ denote the set of real $N \times 1$ vectors and set of real $M \times N$ matrices, respectively; $\mathbf{1}_N$, $\mathbf{I}_N$, and $\mathbf{e}_i$ represent the $N \times 1$ all-one vector, $N \times N$ identity matrix, and unit column vector with the ith entry equal to 1, respectively; "$\succeq$," "$\|\cdot\|$," and "$\backslash$" stand for the componentwise inequality, Euclidean norm, and set difference, respectively; $\text{rank}(\mathbf{X})$, $\det(\mathbf{X})$, $\|\mathbf{X}\|_F$, and $\mathbf{X}^\dagger$ denote the rank, determinant, Frobenius norm, and pseudoinverse of the matrix $\mathbf{X}$, respectively; $[\mathbf{x}]_i$ and $[\mathbf{x}]_{1:i}$ denote the ith element of $\mathbf{x}$ and an $i \times 1$ column vector formed by the first i elements in $\mathbf{x}$, respectively; $\{\mathbf{x}[n]\}_{n=1}^L$ denotes $\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}$; $[\mathbf{x}[n]]_{n \in \mathcal{I}}$ represents a matrix comprising $\mathbf{x}[n]$ for all n's in the set $\mathcal{I}$ as its column vectors; and $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ denotes the Gaussian distribution with zero mean and covariance matrix $\sigma^2 \mathbf{I}_M$.

II. HYPERSPECTRAL ENDMEMBER EXTRACTION PROBLEM

Consider a scenario where a hyperspectral sensor measures solar EM radiation over M spectral bands from N distinct substances in an area of interest. Assuming that the EM patterns are received via only one single reflection and that the materials therein are distinct, each observed hyperspectral pixel vector can be represented by a linear mixing of these substance spectra [1], [2]

$$\mathbf{y}[n] = \mathbf{x}[n] + \mathbf{w}[n] + \mathbf{z}[n], \quad n = 1, \ldots, L, \qquad (1)$$

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] = \sum_{i=1}^N s_i[n]\mathbf{a}_i, \quad n = 1, \ldots, L. \qquad (2)$$

In (1), $\mathbf{y}[n] = [y_1[n], \ldots, y_M[n]]^T$ is the nth noise- and outlier-contaminated pixel vector that comprises M spectral bands, $\mathbf{x}[n] = [x_1[n], \ldots, x_M[n]]^T$ is the contamination-free counterpart, $\mathbf{w}[n] = [w_1[n], \ldots, w_M[n]]^T$ is the isotropically distributed noise vector, e.g., $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ where σ^2 is the noise variance, $\mathbf{z}[n] = [z_1[n], \ldots, z_M[n]]^T$ denotes the outlier vector appearing only at Z pixels, i.e.,

$$\begin{aligned} \mathbf{z}[n] &\neq \mathbf{0}, \quad n \in \{\ell_1, \ldots, \ell_Z\} \triangleq \mathcal{I} \\ \mathbf{z}[n] &= \mathbf{0}, \quad n \in \{1, \ldots, L\} \backslash \mathcal{I} \end{aligned} \qquad (3)$$

and L is the total number of observed pixel vectors. In (2), $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \in \mathbb{R}^{M \times N}$ represents the endmember signature matrix whose ith column vector $\mathbf{a}_i$ denotes the ith endmember signature and $\mathbf{s}[n] = [s_1[n], \ldots, s_N[n]]^T$ is the nth abundance vector comprising N fractional abundances.

In this paper, as mentioned in the introduction section, we focus on the so-called "dead" or "bad" outlier pixels that can be

modeled as $\mathbf{z}[n]$, $n \in \mathcal{I}$, in $\mathbf{y}[n]$ given by (1), and these outliers $\mathbf{z}[n]$, $n \in \mathcal{I}$, are considered without assuming any statistical priors. The outlier pixels are assumed to be rare, and hence, the number of outlier pixels Z should be much less than the number of data samples L, i.e., outlier-data amount ratio $Z/L \ll 1$. While our emphasis is placed on the outliers defined as the dead pixels, the outliers given by (3) can also be interpreted as errors of linear approximation to the real hyperspectral data where the nonlinear mixing model could be a better fit [2], [27].

Assuming prior knowledge of the number of endmembers N, robust hyperspectral endmember extraction is to robustly estimate the endmember signatures $\mathbf{a}_1, \ldots, \mathbf{a}_N$ from the given corrupted hyperspectral data $\mathbf{y}[1], \ldots, \mathbf{y}[L]$ with minimum effects of noise and outliers. Some general assumptions [1], [2] are as follows.

(A1) The intensities of all the abundance vectors are nonnegative, i.e., $s_i[n] \geq 0$ for all i and n.

(A2) Abundance fractions are proportionally distributed for each $\mathbf{s}[n]$, i.e., $\sum_{i=1}^{N} s_i[n] = 1$, $\forall n$.

(A3) $\min\{L, M\} \geq N$ and the endmember signatures $\mathbf{a}_1, \ldots, \mathbf{a}_N$ are linearly independent, i.e., $\mathrm{rank}(\mathbf{A}) = N$.

(A4) (Pure pixel assumption) There exists at least a set of indices $\{l_1, l_2, \ldots, l_N\}$ such that $\mathbf{x}[l_i] = \mathbf{a}_i$ for $i = 1, \ldots, N$, and the set of pure pixels $\{l_1, l_2, \ldots, l_N\}$ and the set of outliers $\mathcal{I}$ are disjoint.

Assumptions (A1), (A2), and (A3) have been widely used in hyperspectral endmember extraction [1], [2], [4], [5]. The pure pixel assumption (A4) fits well for a scenario where the sensor flies in low altitude [28].

It should be noted that the estimation of the number of endmembers N is generally treated as a separate topic [23], [29]–[33]. In this paper, if there are rare (or target) endmembers present in the data, the number of endmembers N should include both background endmembers and rare endmembers. However, when one is only concerned about the background endmembers, then N can be set to the number of background endmembers. Subsequently, the rare endmembers are treated as outliers. Relevant discussions and simulations will be presented in Remark 2 (in Section III) and Section VI-G, respectively.

III. RASF FOR DIMENSION REDUCTION

Dimension reduction is a common primary step for hyperspectral image analysis with the prime merit of reducing the noise effect and computational complexity of the subsequent endmember extraction. In [17], we have made use of affine data geometry for dimension reduction, where it was shown that the affine set estimate provides the best representation to the given hyperspectral data in the least squares error sense. Nevertheless, in practical scenarios where the outliers are present, the fitted affine set could be severely affected by the outliers. To this end, we herein propose a RASF algorithm, attempting to provide an affine set estimate robust against both noise and outliers.

To start with, let us consider the contamination-free pixels $\mathbf{x}[n]$. It has been shown in [17] that, by (A2), the affine hull of $\mathbf{x}[n]$ is identical to that of endmembers $\mathbf{a}_1, \ldots, \mathbf{a}_N$

$$\mathrm{aff}\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\} = \mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \quad (4)$$

where $\mathrm{aff}\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ denotes the affine hull of $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and it is defined as [34]

$$\mathrm{aff}\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} = \left\{ \sum_{i=1}^{N} \theta_i \mathbf{x}_i \,\middle|\, \mathbf{1}_N^T \boldsymbol{\theta} = 1, \boldsymbol{\theta} \in \mathbb{R}^N \right\} \quad (5)$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T$. By (A3), the endmember affine hull $\mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ admits an affine set representation

$$\mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} = \{\mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \,|\, \boldsymbol{\alpha} \in \mathbb{R}^{N-1}\} \triangleq \mathcal{A}(\mathbf{C}, \mathbf{d}) \quad (6)$$

for some (nonunique) affine set parameter $(\mathbf{C}, \mathbf{d}) \in \mathbb{R}^{M \times (N-1)} \times \mathbb{R}^M$ and $\mathrm{rank}(\mathbf{C}) = N - 1$. By virtue of (4) and (6), the dimension reduction of the contamination-free data $\mathbf{x}[n]$ can be easily carried out by

$$\tilde{\mathbf{x}}[n] = \mathbf{C}^{\dagger}(\mathbf{x}[n] - \mathbf{d}) = \sum_{i=1}^{N} s_i[n]\boldsymbol{\alpha}_i, \quad n = 1, \ldots, L \quad (7)$$

where

$$\boldsymbol{\alpha}_i = \mathbf{C}^{\dagger}(\mathbf{a}_i - \mathbf{d}), \quad i = 1, \ldots, N \quad (8)$$

are the dimension-reduced endmembers. There exists a closed-form solution to the affine set parameter $(\mathbf{C}, \mathbf{d})$ if the acquisition of the $\{\mathbf{x}[n]\}_{n=1}^{L}$ is possible [17]. However, what we have in reality is the contaminated observed pixel vectors $\{\mathbf{y}[n]\}_{n=1}^{L}$, and therefore, obtaining an accurate estimate of $(\mathbf{C}, \mathbf{d})$ from $\{\mathbf{y}[n]\}_{n=1}^{L}$ will be a challenging problem.

To tackle this issue, we consider the following RASF problem:

$$\min_{\mathrm{num}\{\mathbf{z}_1, \ldots, \mathbf{z}_L\} \leq Z} \left\{ \min_{\substack{\mathbf{x}_n \in \mathcal{A}(\mathbf{C}, \mathbf{d}) \\ \mathbf{C}^T \mathbf{C} = \mathbf{I}_{N-1} \\ n=1, \ldots, L}} \sum_{n=1}^{L} \|\mathbf{y}[n] - \mathbf{x}_n - \mathbf{z}_n\|^2 \right\} \quad (9)$$

where $\mathrm{num}\{\mathbf{z}_1, \ldots, \mathbf{z}_L\}$ denotes the number of nonzero vectors in $\{\mathbf{z}_1, \ldots, \mathbf{z}_L\}$ and the number of outliers Z is assumed to be known for ease of our derivations. Some discussions on how we set a value of Z are given in Remark 1 hereinafter. The objective of (9) is to seek an $(N-1)$-dimensional affine set $\mathcal{A}(\mathbf{C}, \mathbf{d})$ with the minimum projection error with respect to (w.r.t.) $\mathbf{y}[n]$ and with the minimum effect of outliers $\mathbf{z}[n]$. Problem (9) is difficult to solve in a globally optimal sense but can be approximated by alternating optimization. Let us consider the following two partial minimization problems.

1) Problem (9) w.r.t. Variables $\{\mathbf{x}_n\}_{n=1}^{L}$, $\mathbf{C}$, and $\mathbf{d}$:

$$\min_{\substack{\mathbf{x}_n \in \mathcal{A}(\mathbf{C}, \mathbf{d}), \mathbf{C}^T \mathbf{C} = \mathbf{I}_{N-1} \\ n=1, \ldots, L}} \sum_{n=1}^{L} \|(\mathbf{y}[n] - \hat{\mathbf{z}}_n) - \mathbf{x}_n\|^2 \quad (10)$$

for any given $\{\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_L\}$ that satisfies $\mathrm{num}\{\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_L\} \leq Z$. Following the proof in [22, Proposition 1], problem (10) can be shown to have an analytical solution given by

$$\hat{\mathbf{d}} = \frac{1}{L} \sum_{n=1}^{L} (\mathbf{y}[n] - \hat{\mathbf{z}}_n) \quad (11)$$

$$\hat{\mathbf{C}} = [\mathbf{q}_1(\mathbf{U}\mathbf{U}^T), \mathbf{q}_2(\mathbf{U}\mathbf{U}^T), \ldots, \mathbf{q}_{N-1}(\mathbf{U}\mathbf{U}^T)] \quad (12)$$

Given a convergence tolerance $\varepsilon > 0$, hyperspectral data $\{\mathbf{y}[n]\}_{n=1}^{L}$, and the number of endmembers N.

Step 1. initialize $\hat{\mathbf{z}}_1 = \cdots = \hat{\mathbf{z}}_L = \mathbf{0}$, and iteration number $k := 0$.

Step 2. update the solution of problem (10)

$$\hat{\mathbf{d}} = \frac{1}{L} \sum_{n=1}^{L} \mathbf{y}[n] - \hat{\mathbf{z}}_n,$$

$$\hat{\mathbf{C}} = [\mathbf{q}_1(\mathbf{U}\mathbf{U}^T), \mathbf{q}_2(\mathbf{U}\mathbf{U}^T), ..., \mathbf{q}_{N-1}(\mathbf{U}\mathbf{U}^T)],$$

$$\hat{\mathbf{x}}_n = \hat{\mathbf{C}}\hat{\mathbf{C}}^T(\mathbf{y}[n] - \hat{\mathbf{z}}_n - \hat{\mathbf{d}}) + \hat{\mathbf{d}}, \ n = 1, ..., L,$$

where $\mathbf{U} = [(\mathbf{y}[1] - \hat{\mathbf{z}}_1) - \hat{\mathbf{d}}, ..., (\mathbf{y}[L] - \hat{\mathbf{z}}_L) - \hat{\mathbf{d}}]$.

Step 3. update the solution of problem (14)

$$\hat{\mathbf{z}}_n = \begin{cases} \mathbf{y}[n] - \hat{\mathbf{x}}_n, & n \in \{\hat{\ell}_1, ..., \hat{\ell}_Z\} \\ \mathbf{0}, & n \in \{1, ..., L\} \setminus \{\hat{\ell}_1, ..., \hat{\ell}_Z\} \end{cases}$$

where $\hat{\ell}_i$ is the index of the ith largest value in $(\|\mathbf{y}[1] - \hat{\mathbf{x}}_1\|, ..., \|\mathbf{y}[L] - \hat{\mathbf{x}}_L\|)$.

Step 4. update $k := k + 1$ and the k-iterate objective value $\varrho(k) = \sum_{n=1}^{L} \|\mathbf{y}[n] - \hat{\mathbf{x}}_n - \hat{\mathbf{z}}_n\|^2$.

Step 5. if $k = 1$ or $(\varrho(k-1) - \varrho(k))/\varrho(k-1) < \varepsilon$ for $k > 1$, then go to **Step 2**,

Step 6. output the approximate robust affine set parameter $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ and the outlier pixel indices $\hat{\mathcal{I}} = \{\hat{\ell}_1, ..., \hat{\ell}_Z\}$.

where $\mathbf{U} = [(\mathbf{y}[1] - \hat{\mathbf{z}}_1) - \hat{\mathbf{d}}, \ldots, (\mathbf{y}[L] - \hat{\mathbf{z}}_L) - \hat{\mathbf{d}}]$ and $\mathbf{q}_i(\mathbf{U}\mathbf{U}^T)$ denotes the unit-norm eigenvector associated with the ith principal eigenvalue of $\mathbf{U}\mathbf{U}^T$. The estimated affinely projected data $\hat{\mathbf{x}}_n \in \mathcal{A}(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ can be easily shown to be

$$\hat{\mathbf{x}}_n = \hat{\mathbf{C}}\hat{\mathbf{C}}^T\left(\mathbf{y}[n] - \hat{\mathbf{z}}_n - \hat{\mathbf{d}}\right) + \hat{\mathbf{d}}, \quad n = 1, \ldots, L. \quad (13)$$

2) Problem (9) w.r.t. Variables $\{\mathbf{z}_n\}_{n=1}^{L}$:

$$\min_{\text{num}\{\mathbf{z}_1, ..., \mathbf{z}_L\} \leq Z} \sum_{n=1}^{L} \|(\mathbf{y}[n] - \hat{\mathbf{x}}_n) - \mathbf{z}_n\|^2 \quad (14)$$

for any given $\{\hat{\mathbf{x}}_n\}_{n=1}^{L} \subset \mathcal{A}(\hat{\mathbf{C}}, \hat{\mathbf{d}})$. It is trivial to see that the solution of the aforementioned problem is

$$\hat{\mathbf{z}}_n = \begin{cases} \mathbf{y}[n] - \hat{\mathbf{x}}_n, & n \in \{\hat{\ell}_1, \ldots, \hat{\ell}_Z\} \\ \mathbf{0}, & n \in \{1, \ldots, L\} \setminus \{\hat{\ell}_1, \ldots, \hat{\ell}_Z\} \end{cases} \quad (15)$$

where $\hat{\ell}_i$ is the index of the ith largest value in $(\|\mathbf{y}[1] - \hat{\mathbf{x}}_1\|, \ldots, \|\mathbf{y}[L] - \hat{\mathbf{x}}_L\|)$.

A solution of problem (9) can be obtained by handling the aforementioned two subproblems in a cyclic manner until some stopping criterion is met. The pseudocodes of the RASF algorithm for (9) are given in Table I.

Similar to (7), the affine set parameter estimate $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ can be used to obtain the dimension-reduced observed pixel vectors

$$\tilde{\mathbf{y}}[n] \triangleq \hat{\mathbf{C}}^T\left(\mathbf{y}[n] - \hat{\mathbf{d}}\right) \cong \tilde{\mathbf{x}}[n] + \hat{\mathbf{C}}^T\mathbf{w}[n] + \hat{\mathbf{C}}^T\mathbf{z}[n] \quad (16)$$

$$= \sum_{i=1}^{N} s_i[n]\boldsymbol{\alpha}_i + \tilde{\mathbf{w}}[n] + \tilde{\mathbf{z}}[n] \in \mathbb{R}^{N-1}, \forall n \quad (17)$$

where "$\cong$" is in (16) because $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ is an approximation to the true $(\mathbf{C}, \mathbf{d})$

$$\tilde{\mathbf{w}}[n] \triangleq \hat{\mathbf{C}}^T\mathbf{w}[n] \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{N-1}) \quad (18)$$

due to $\mathbf{w}[n] \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ and $\hat{\mathbf{C}}^T\hat{\mathbf{C}} = \mathbf{I}_{N-1}$, and $\tilde{\mathbf{z}}[n] \triangleq \hat{\mathbf{C}}^T\mathbf{z}[n]$. By (A1) and (A2), the dimension-reduced contamination-free data $\tilde{\mathbf{x}}[n]$ given by (7) must be in the convex hull of $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ [34], denoted by

$$\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} = \left\{\sum_{i=1}^{N} \theta_i\boldsymbol{\alpha}_i \middle| \boldsymbol{\theta} \succeq \mathbf{0}, \mathbf{1}_N^T\boldsymbol{\theta} = 1\right\} \quad (19)$$

but the contaminations caused by the noise $\tilde{\mathbf{w}}[n]$ and outlier $\tilde{\mathbf{z}}[n]$ could possibly make the observed pixel $\tilde{\mathbf{y}}[n]$ given by (17) out of the $\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$. Fig. 1(a) and (b) illustrate the geometries of the original data $\{\mathbf{y}[n]\}_{n=1}^{L}$ and the dimension-reduced data $\{\tilde{\mathbf{y}}[n]\}_{n=1}^{L}$, respectively, for $N = 3$ and $Z = 5$.

As reported in [25] that the outlier pixels could significantly affect the results of subsequent endmember extraction, we remove the outlier pixels at $\hat{\mathcal{I}} = \{\hat{\ell}_1, \ldots, \hat{\ell}_Z\}$ from the data $\{\tilde{\mathbf{y}}[n]\}_{n=1}^{L}$, and hence, the endmember extraction problem is then to estimate $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$ from

$$\tilde{\mathbf{y}}[n] \cong \sum_{i=1}^{N} s_i[n]\boldsymbol{\alpha}_i + \tilde{\mathbf{w}}[n], n \in \{1, \ldots, L\} \setminus \hat{\mathcal{I}} \quad (20)$$

where the noise $\tilde{\mathbf{w}}[n]$ is still present. The robust EEAs to be proposed in Section IV will take the noise effect into account. Once $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$ are obtained, one can simply recover the endmember estimates by the following affine transformation:

$$\hat{\mathbf{a}}_i = \hat{\mathbf{C}}\boldsymbol{\alpha}_i + \hat{\mathbf{d}}, \quad i = 1, \ldots, N. \quad (21)$$

Let us conclude this section with two remarks on the choice of Z and N for the proposed RASF algorithm in practical scenarios.

Remark 1: The number of outliers Z is impossible to be known *a priori* in practical scenarios, but as will be seen in our simulations, the solution $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ obtained by the RASF algorithm in Table I is insensitive to the preset number of outliers, denoted by $\hat{Z}$, when $\hat{Z} \geq Z$, meaning that the estimated $\hat{Z}$ outliers sufficiently cover the Z true outliers. However, setting $\hat{Z}$ too large for the RASF algorithm could be jeopardous—it may lead the RASF algorithm to incidently mistake some rare endmembers as outliers. How we practically select the value of Z in real applications will be considered as our future direction.

Remark 2: In a scenario where the N endmembers include both background and rare endmembers, it will be seen from our simulations that the RASF algorithm can preserve the rare endmembers in the subsequent endmember extraction process, provided that the N is perfectly estimated. It is also suggested by the simulations that, if one cares only about background endmembers without the desire of preserving rare endmembers, in the proposed RASF algorithm, the number of endmembers can be set to the number of background endmembers only. Then, the proposed RASF will automatically find $(\mathbf{C}, \mathbf{d})$ with the minimum impact of both "dead" pixels and rare endmembers. How to estimate N in the presence of rare endmembers can be referred to [31] and [32].
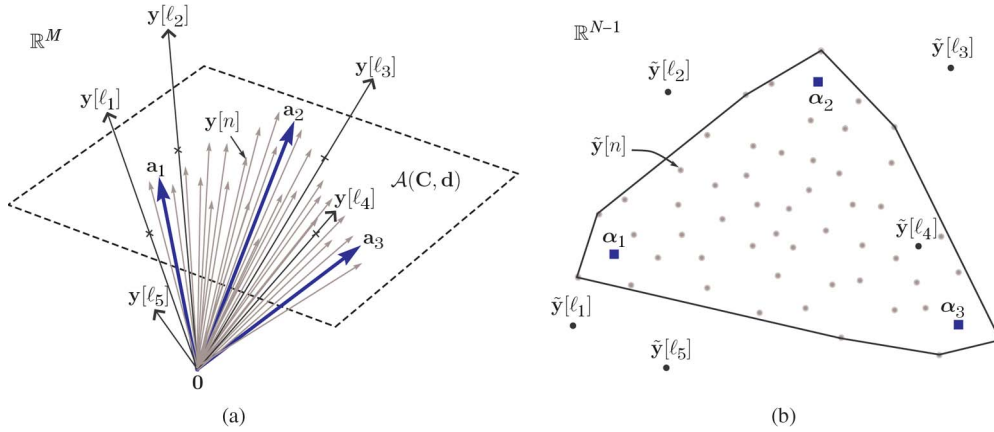
Fig. 1.   Illustration of (a) the original data $\{\mathbf{y}[n]\}_{n=1}^{L}$ and (b) the dimension-reduced data $\{\tilde{\mathbf{y}}[n]\}_{n=1}^{L}$ for $N = 3$ and $Z = 5$.

IV. REVIEW OF WINTER'S ENDMEMBER EXTRACTION AND ITS ROBUST GENERALIZATION

In the past decade, Winter's maximum-volume simplex criterion has led to the well-known EEA, N-FINDR, and its many variants [21], [19], [20], [35], [36]. Suppose that the outlier pixels have been perfectly identified and removed from the data by the RASF algorithm, i.e., $\{\tilde{\mathbf{y}}[n]\}_{n\in\{1,...,L\}\backslash\hat{\mathcal{I}}}$ and $\hat{\mathcal{I}} = \mathcal{I}$. N-FINDR is to estimate endmembers by finding the vertices of the maximum-volume simplex inside the dimension-reduced data cloud $\{\tilde{\mathbf{y}}[n]\}_{n\in\{1,...,L\}\backslash\mathcal{I}}$. Recently, we revisited the Winter's endmember extraction criterion and reported two N-FINDR variants in [21], based on the following continuous formulation of Winter's criterion:

$$\max_{\mathbf{v}_1,...,\mathbf{v}_N\in\mathbb{R}^{N-1}} \mathrm{vol}(\mathbf{v}_1,...,\mathbf{v}_N)$$

$$\text{s.t.} \quad \mathbf{v}_i \in \mathrm{conv}\left\{\{\tilde{\mathbf{y}}[n]\}_{n\in\{1,...,L\}\backslash\mathcal{I}}\right\}, \ \forall i \quad (22)$$

where

$$\mathrm{vol}(\mathbf{v}_1,...,\mathbf{v}_N) = \frac{|\det(\mathbf{\Delta}(\mathbf{v}_1,...,\mathbf{v}_N))|}{(N-1)!} \quad (23)$$

is the volume of $\mathrm{conv}\{\mathbf{v}_1,...,\mathbf{v}_N\}$, and

$$\mathbf{\Delta}(\mathbf{v}_1,...,\mathbf{v}_N) = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \\ 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{N\times N}. \quad (24)$$

It has been theoretically proved that the true endmembers can be perfectly estimated by solving problem (22) under (A1) to (A4) and in the absence of noise, i.e., $\tilde{\mathbf{y}}[n] = \tilde{\mathbf{x}}[n]$, $\forall n$ [21]. However, in the presence of additive isotropic random noise, the simplex volume yielded by Winter's criterion may be larger than that of the true simplex [21]. In other words, the endmember estimates obtained by Winter's criterion may be far away from the true endmembers when the observed data are corrupted by noise. To mitigate such simplex inflation, in [21], we reported an idea to pull back Winter's endmember estimates by a suitable margin such that $(\boldsymbol{\nu}_1,...,\boldsymbol{\nu}_N)$ are closer to the true endmembers $(\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_N)$. This idea, as illustrated in
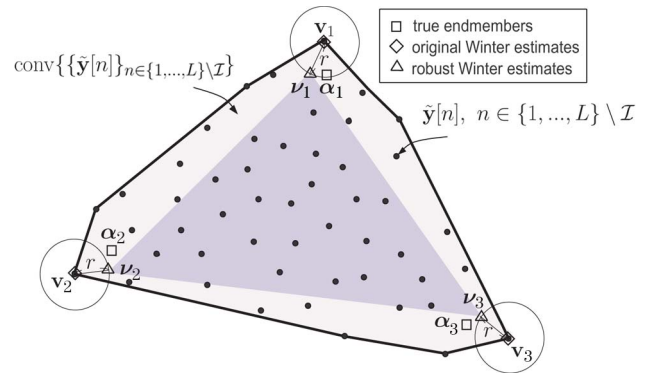

Fig. 2.   Illustration of robust Winter's endmember extraction problem for $N = 3$.

Fig. 2, can be formulated as the following simplex volume max-min problem [21]:

$$\max_{\substack{\mathbf{v}_i\in\mathbb{R}^{N-1}, \\ i=1,...,N}} \left\{ \min_{\substack{\|\mathbf{u}_i\|\leq r, \\ i=1,...,N}} \left| \det(\mathbf{\Delta}(\mathbf{v}_1-\mathbf{u}_1,...,\mathbf{v}_N-\mathbf{u}_N)) \right| \right\}$$

$$\text{s.t.} \quad \mathbf{v}_i \in \mathrm{conv}\left\{\{\tilde{\mathbf{y}}[n]\}_{n\in\{1,...,L\}\backslash\mathcal{I}}\right\}, \ \forall i=1,...,N \quad (25)$$

where each $\mathbf{u}_i$ lying in a norm ball $\{\mathbf{u}\in\mathbb{R}^{N-1}| \ \|\mathbf{u}\|\leq r\}$ denotes the pull-back vector and r is the maximum back-off distance. Denoting the optimal solution of problem (25) by $(\hat{\mathbf{v}}_1,...,\hat{\mathbf{v}}_N,\hat{\mathbf{u}}_1,...,\hat{\mathbf{u}}_N)$, the *robust endmember estimates* are obtained as

$$\hat{\boldsymbol{\nu}}_i = \hat{\mathbf{v}}_i - \hat{\mathbf{u}}_i, \quad i=1,...,N. \quad (26)$$

In [21], we have proposed an algorithm for handling problem (25), called WAVMAX. WAVMAX has demonstrated performance improvement in the noisy scenario, but it is computationally expensive. We therefore propose two fast algorithms to handle problem (25) in the next section.

V. FAST ALGORITHMS FOR SIMPLEX VOLUME MAX-MIN PROBLEM

In this section, two fast algorithms are proposed for handling the simplex volume max-min formulation (25). We employ a

decoupled max-min heuristic, and different partial optimization schemes to handle problem (25), and come up with two algorithms, called ADVMM and SDVMM, respectively. The computational complexity orders of the proposed methods are also discussed.

A. ADVMM Algorithm

We first reformulate problem (25) for ease of development. Denote the outlier-free data matrix by $\widetilde{\mathbf{Y}} = [\tilde{\mathbf{y}}[n]]_{n \in \{1,\dots,L\}\setminus\mathcal{I}} \in \mathbb{R}^{(N-1)\times(L-Z)}$. Then, by the convex combination expression

$$\mathbf{v}_i = \widetilde{\mathbf{Y}}\boldsymbol{\theta}_i \tag{27}$$

and the property $\det(\mathbf{P}\boldsymbol{\Delta}) = \pm\det(\boldsymbol{\Delta})$ for any permutation matrix $\mathbf{P}$, problem (25) can be expressed as

$$\max_{\substack{\boldsymbol{\theta}_i \in \mathcal{S}, \\ i=1,\dots,N}} \left\{ \min_{\substack{\|\mathbf{u}_i\| \leq r, \\ i=1,\dots,N}} \det\left(\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\boldsymbol{\theta}_1 - \mathbf{u}_1, \dots, \widetilde{\mathbf{Y}}\boldsymbol{\theta}_N - \mathbf{u}_N)\right) \right\} \tag{28}$$

where $\mathcal{S} = \{\boldsymbol{\theta} \in \mathbb{R}^{L-Z} | \boldsymbol{\theta} \succeq \mathbf{0}, \mathbf{1}_{L-Z}^T\boldsymbol{\theta} = 1\}$. Optimizing $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N$ and $\mathbf{u}_1, \dots, \mathbf{u}_N$ jointly in (28) is quite challenging. In ADVMM, we consider the partial max-min problem of (28) w.r.t. the pair $(\boldsymbol{\theta}_j, \mathbf{u}_j)$ while fixing the other pairs $(\boldsymbol{\theta}_i, \mathbf{u}_i)$ for $i \neq j$; such partial max-min problems are represented by

$$\max_{\boldsymbol{\theta}_j \in \mathcal{S}} \left\{ \min_{\|\mathbf{u}_j\| \leq r} \det\left(\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{u}}_1, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_{j-1} - \hat{\mathbf{u}}_{j-1}, \right.\right.$$
$$\left.\left. \widetilde{\mathbf{Y}}\boldsymbol{\theta}_j - \mathbf{u}_j, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_{j+1} - \hat{\mathbf{u}}_{j+1}, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_N - \hat{\mathbf{u}}_N)\right) \right\}. \tag{29}$$

The partial max-min problems (29) for $j = 1, \dots, N$ are conducted cyclically until some stopping criterion is satisfied. A connection of the aforementioned decoupled max-min problem to the original alternating maximization of (28) used in WAVMAX will be discussed in the end of this section; see Remark 3.

Next, we will present how to solve the partial max-min problem (29). By applying a cofactor expansion of $\det(\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{u}}_1, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_N - \hat{\mathbf{u}}_N))$ along the jth column, we have

$$\det\left(\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{u}}_1, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_N - \hat{\mathbf{u}}_N)\right)$$
$$= \mathbf{k}_j^T(\widetilde{\mathbf{Y}}\boldsymbol{\theta}_j - \mathbf{u}_j) + (-1)^{N+j}\det(\boldsymbol{\mathcal{Q}}_{Nj}) \tag{30}$$

where $\mathbf{k}_j \in \mathbb{R}^{N-1}$ is expressed as

$$\mathbf{k}_j = \left[(-1)^{1+j}\det(\boldsymbol{\mathcal{Q}}_{1j}), \dots, (-1)^{N-1+j}\det\left(\boldsymbol{\mathcal{Q}}_{(N-1)j}\right)\right]^T \tag{31}$$

and $\boldsymbol{\mathcal{Q}}_{ij} \in \mathbb{R}^{(N-1)\times(N-1)}$ is a submatrix of $\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{u}}_1, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_N - \hat{\mathbf{u}}_N)$ with the ith row and the jth column removed. Then, problem (29) is equivalent to

$$\max_{\boldsymbol{\theta}_j \in \mathcal{S}} \left\{ \min_{\|\mathbf{u}_j\| \leq r} \mathbf{k}_j^T(\widetilde{\mathbf{Y}}\boldsymbol{\theta}_j - \mathbf{u}_j) \right\} \tag{32}$$

where the term $(-1)^{N+j}\det(\boldsymbol{\mathcal{Q}}_{Nj})$ in (30) is independent of $(\boldsymbol{\theta}_j, \mathbf{u}_j)$ and so is removed without loss of optimality. In addition, since $\boldsymbol{\theta}_j$ and $\mathbf{u}_j$ have been decoupled, problem (32)

Given	a convergence tolerance $\varepsilon > 0$, a back-off distance r, the dimension reduced and outlier-removed data matrix $\widetilde{\mathbf{Y}}$ and the number of endmembers N.
Step 1.	randomly select $(\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_N)$ from $\{\mathbf{e}_i\}_{i=1}^{L-Z}$ and set $\hat{\mathbf{u}}_1 = \dots = \hat{\mathbf{u}}_N = \mathbf{0}$.
Step 2.	set $j := 1$, $\varrho := \det(\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{u}}_1, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_N - \hat{\mathbf{u}}_N))$.
Step 3.	compute $\mathbf{k}_j$ by (31), and then update $\hat{\mathbf{u}}_j$ by (33) and $\hat{\boldsymbol{\theta}}_j$ by (34).
Step 4.	if $(j \text{ modulo } N) \neq 0$, then $j := j + 1$ and go to **Step 3**, else compute $\bar{\varrho} = \det(\boldsymbol{\Delta}(\widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_1 - \hat{\mathbf{u}}_1, \dots, \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_N - \hat{\mathbf{u}}_N))$.
Step 5.	if $|\bar{\varrho} - \varrho|/\varrho > \varepsilon$, then set $\varrho := \bar{\varrho}$, $j := 1$, and go to **Step 3**, else output $\hat{\boldsymbol{\nu}}_j = \widetilde{\mathbf{Y}}\hat{\boldsymbol{\theta}}_j - \hat{\mathbf{u}}_j$, $\forall j$ as an approximate solution to (25).

can be handled by solving the following two separate problems, each with a closed-form solution:

$$\hat{\mathbf{u}}_j = \arg\max_{\|\mathbf{u}_j\| \leq r} \mathbf{k}_j^T\mathbf{u}_j = \frac{r\mathbf{k}_j}{\|\mathbf{k}_j\|} \tag{33}$$

$$\hat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta}_j \in \mathcal{S}} \mathbf{k}_j^T\widetilde{\mathbf{Y}}\boldsymbol{\theta}_j = \mathbf{e}_l, \quad l = \arg\max_{n \in \{1,\dots,L\}\setminus\mathcal{I}} \mathbf{k}_j^T\tilde{\mathbf{y}}[n] \tag{34}$$

where $\hat{\mathbf{u}}_j$ in (33) is obtained by the Cauchy–Schwarz inequality and $\hat{\boldsymbol{\theta}}_j$ in (34) can be obtained by [21, Lemma 2]. Let us summarize how we pragmatically implement the ADVMM algorithm in Table II.

Remark 3: WAVMAX, as has been presented in [21], is an alternating optimization method w.r.t. $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N$. Its jth alternating maximization problem to (28) can be expressed as

$$\max_{\boldsymbol{\theta}_j \in \mathcal{S}} \min_{\substack{\|\mathbf{u}_i\| \leq r, \\ i=1,\dots,N}} f(\boldsymbol{\theta}_j, \hat{\boldsymbol{\Theta}}_j, \mathbf{u}_1, \dots, \mathbf{u}_N) \tag{35}$$

where $f(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N, \mathbf{u}_1, \dots, \mathbf{u}_N)$ is the objective function of (28) and $\hat{\boldsymbol{\Theta}}_j = [\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_{j-1}, \hat{\boldsymbol{\theta}}_{j+1}, \dots, \hat{\boldsymbol{\theta}}_N]$ is fixed. As can be observed from (35), for each update $\boldsymbol{\theta}_j$, we have to deal with the inner minimization $\min f(\boldsymbol{\theta}_j, \hat{\boldsymbol{\Theta}}_j, \mathbf{u}_1, \dots, \mathbf{u}_N)$ w.r.t. $\mathbf{u}_1, \dots, \mathbf{u}_N$ jointly. In [21], we have used the subgradient method to handle (35), but the resulting WAVMAX algorithm is computationally complicated. The proposed ADVMM algorithm uses a computationally efficient way to approximate (35), which, in turn, handles the jth decoupled max-min problem (29), or equivalently

$$\max_{\boldsymbol{\theta}_j \in \mathcal{S}} \min_{\|\mathbf{u}_j\| \leq r} f(\boldsymbol{\theta}_j, \hat{\boldsymbol{\Theta}}_j, \mathbf{u}_j, \hat{\mathbf{U}}_j) \tag{36}$$

where $\hat{\mathbf{U}}_j = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_{j-1}, \hat{\mathbf{u}}_{j+1}, \dots, \hat{\mathbf{u}}_N]$. It is easy to observe that problem (36) serves as an upper bound of problem (35). If $\hat{\mathbf{U}}_j$ happens to be the optimal solution of the inner minimization of (35), then the upper bound (36) will be equal to (35). As a result, the proposed ADVMM algorithm can be thought of as a method to maximize an upper bound of the partial maximization problem of (28) in WAVMAX.

B. SDVMM Algorithm

We turn our attention to how we apply successive optimization to handle problem (25). By letting

$$\mathbf{w}_i = \begin{bmatrix}\mathbf{v}_i^T & 1\end{bmatrix}^T, \quad \mathbf{t}_i = \begin{bmatrix}\mathbf{u}_i^T & 0\end{bmatrix}^T, \quad \bar{\mathbf{y}}[n] = \begin{bmatrix}\tilde{\mathbf{y}}[n]^T & 1\end{bmatrix}^T \tag{37}$$

problem (25) can be rewritten as

$$
\max_{\substack{\mathbf{w}_i \in \mathcal{F}, \\ i=1,\dots,N}} \left\{ \min_{\substack{\|\mathbf{t}_i\| \le r, \\ \mathbf{e}_N^T \mathbf{t}_i = 0, \forall i}} \left| \det\left([\mathbf{w}_1 - \mathbf{t}_1, \dots, \mathbf{w}_N - \mathbf{t}_N] \right) \right| \right\} \quad (38)
$$

where $\mathcal{F} = \operatorname{conv}\{\{\bar{\mathbf{y}}[n]\}_{n \in \{1,\dots,L\} \setminus \mathcal{I}}\}$. It has been shown in [21, Lemma 3] that

$$
\left| \det\left([\mathbf{w}_1 - \mathbf{t}_1, \dots, \mathbf{w}_N - \mathbf{t}_N] \right) \right|
$$
$$
= \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:0}} (\mathbf{w}_1 - \mathbf{t}_1) \right\| \dots \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(N-1)}} (\mathbf{w}_N - \mathbf{t}_N) \right\| \quad (39)
$$

where $\mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:j}} = \mathbf{I}_N - \mathbf{H}_{1:j}(\mathbf{H}_{1:j}^T \mathbf{H}_{1:j})^\dagger \mathbf{H}_{1:j}^T$ is the orthogonal complement projector of

$$
\mathbf{H}_{1:j} \triangleq [\mathbf{w}_1 - \mathbf{t}_1, \dots, \mathbf{w}_j - \mathbf{t}_j] \quad (40)
$$

and $\mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:0}} = \mathbf{I}_N$. Hence, substituting (39) into problem (38) yields

$$
\max_{\substack{\mathbf{w}_i \in \mathcal{F}, \\ i=1,\dots,N}} \min_{\substack{\|\mathbf{t}_i\| \le r, \\ \mathbf{e}_N^T \mathbf{t}_i = 0, \forall i}} \prod_{j=1}^{N} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) \right\|. \quad (41)
$$

Solving problem (41) w.r.t. $2N$-tuple $(\mathbf{w}_1, \dots, \mathbf{w}_N, \mathbf{t}_1, \dots, \mathbf{t}_N)$ is difficult. In SDVMM, we decouple problem (41) into a set of max-min subproblems and employ successive optimization to these subproblems as follows:

$$
(\hat{\mathbf{w}}_j, \hat{\mathbf{t}}_j) = \arg \max_{\mathbf{w}_j \in \mathcal{F}} \min_{\substack{\|\mathbf{t}_j\| \le r, \\ \mathbf{e}_N^T \mathbf{t}_j = 0}} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) \right\| \quad (42)
$$

from $j = 1$ to N. The solution $(\hat{\mathbf{w}}_j, \hat{\mathbf{t}}_j)$ is obtained by handling the jth max-min subproblem with the previous $(j-1)$ max-min subproblem solutions $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_{j-1}, \hat{\mathbf{t}}_1, \dots, \hat{\mathbf{t}}_{j-1}$ used in $\widehat{\mathbf{H}}_{1:(j-1)}$ as defined in (40). Unlike alternating optimization, the methodology presented here is initialization free and only needs to solve (42) successively for $j = 1, \dots, N$. A relation of the successive optimization procedure given by (42) to problem (41) will be discussed in Remark 4 hereinafter.

The issue that remains is how we handle each difficult (nonconvex) max-min subproblem (42). By relaxing $\mathbf{e}_N^T \mathbf{t}_j = 0$, it can be shown that a closed-form solution to (42) exists. To see this, problem (42) with $\mathbf{e}_N^T \mathbf{t}_j = 0$ relaxed is

$$
\max_{\mathbf{w}_j \in \mathcal{F}} \min_{\|\mathbf{t}_j\| \le r} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) \right\|, \quad j = 1, \dots, N. \quad (43)
$$

The inner problem of (43) for any $\mathbf{w}_j \in \mathcal{F}$ is

$$
\hat{\mathbf{t}}_j = \arg \min_{\|\mathbf{t}_j\| \le r} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) \right\|. \quad (44)
$$

Problem (44) is convex, and Slater's condition holds [34]. The optimal solution of problem (44) can be derived by Karush–Kuhn–Tucker (KKT) conditions, as stated in the following lemma.

TABLE III
SDVMM ALGORITHM FOR PROBLEM (25)

Given	a back-off distance r, the dimension reduced and outlier-removed data set $\{\tilde{\mathbf{y}}[n]\}_{n \in \{1,\dots,L\} \setminus \mathcal{I}}$ and the number of endmembers N.
Step 1.	construct $\bar{\mathbf{y}}[n] = [\tilde{\mathbf{y}}[n]^T \ 1]^T$, $n \in \{1,\dots,L\} \setminus \mathcal{I}$ and set $\widehat{\mathbf{H}}_{1:0} = \mathbf{I}_N$ and $j = 0$.
Step 2.	update $j := j + 1$ and obtain $\hat{\mathbf{w}}_j$ by (48), and $\hat{\mathbf{t}}_j$ by (45).
Step 3.	set $[\hat{\mathbf{t}}_j]_N = 0$, update $\widehat{\mathbf{H}}_{1:j} := [\widehat{\mathbf{H}}_{1:(j-1)} \ \hat{\mathbf{w}}_j - \hat{\mathbf{t}}_j]$ and go to **Step 2** until $j = N$.
Step 4.	output $\hat{\boldsymbol{\nu}}_j = [\hat{\mathbf{w}}_j]_{1:N-1} - [\hat{\mathbf{t}}_j]_{1:N-1}$, $\forall j$ as an approximate solution to (25).

Lemma 1: For any $\mathbf{w}_j \in \mathcal{F}$, problem (44) has an analytical solution given by

$$
\hat{\mathbf{t}}_j = \frac{r \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j}{\left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j \right\|}, \quad \mathbf{w}_j \in \mathcal{W}(r) \quad (45)
$$

$$
\hat{\mathbf{t}}_j \in \left\{ \mathbf{t}_j \mid \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) = \mathbf{0} \right\}, \quad \mathbf{w}_j \in \mathbb{R}^N \setminus \mathcal{W}(r) \quad (46)
$$

where $\mathcal{W}(r) = \{\mathbf{w} \in \mathbb{R}^N \mid \|\mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}\| > r\}$.

Proof: The proof of Lemma 1 is given in the Appendix. ∎

It is trivial to see that the solution (46) always yields zero objective value in (43), and hence, only the optimal solution (45) is considered. Substituting (45) into (43) yields

$$
\max_{\mathbf{w}_j \in \mathcal{F} \bigcap \mathcal{W}(r)} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j \right\|. \quad (47)
$$

The optimal solution of (47) can be easily obtained by following the proof in [21, Lemma 4]; it is given by

$$
\hat{\mathbf{w}}_j = \bar{\mathbf{y}}[l], \quad l = \arg \max_{n \in \mathcal{N}_j} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} \bar{\mathbf{y}}[n] \right\| \quad (48)
$$

where $\mathcal{N}_j = \{n \mid \|\mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} \bar{\mathbf{y}}[n]\| > r, n \in \{1, \dots, L\} \setminus \mathcal{I}\}$.

We should mention that the constraint $\mathbf{w}_j \in \mathcal{W}(r)$ is to ensure the nontrivial solution of problem (43). In fact, one can properly choose an r such that $\mathbf{w}_j \in \mathcal{W}(r)$, $j = 1, \dots, N$, are all satisfied. Also, if the $(\hat{\mathbf{w}}_j, \hat{\mathbf{t}}_j)$ is obtained, we can artificially set $[\hat{\mathbf{t}}_j]_N = 0$ to ensure the feasibility of $(\hat{\mathbf{w}}_j, \hat{\mathbf{t}}_j)$ to problem (42). The pseudocodes of the SDVMM algorithm are given in Table III.

Remark 4: A theoretical justification of the decoupled max-min heuristic used in the proposed SDVMM algorithm is presented herein. By Von Neumann's max-min theorem (or max-min inequality)[1] [37], one can easily derive an upper bound of (41) as follows:

$$
\max_{\substack{\mathbf{w}_N \in \mathcal{F}}} \min_{\substack{\|\mathbf{t}_N\| \le r, \\ \mathbf{e}_N^T \mathbf{t}_N = 0}} \left[\dots \left[\max_{\substack{\mathbf{w}_1 \in \mathcal{F}}} \min_{\substack{\|\mathbf{t}_1\| \le r, \\ \mathbf{e}_N^T \mathbf{t}_1 = 0}} \prod_{j=1}^{N} \left\| \mathbf{P}^\perp_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) \right\| \right] \dots \right]. \quad (49)
$$

Obviously, applying successive optimization to problem (49) turns out to be the same as the optimization procedure described

---

[1] The max-min inequality states that, for any real function $f : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ and any real sets $\mathcal{P} \subseteq \mathbb{R}^N$ and $\mathcal{Q} \subseteq \mathbb{R}^M$, it holds true that $\min_{\mathbf{p} \in \mathcal{P}} \max_{\mathbf{q} \in \mathcal{Q}} f(\mathbf{p}, \mathbf{q}) \ge \max_{\mathbf{q} \in \mathcal{Q}} \min_{\mathbf{p} \in \mathcal{P}} f(\mathbf{p}, \mathbf{q})$.

in (42). Hence, SDVMM algorithm can be thought of as a method that approximates an upper bound of (38), i.e., (49), using successive optimization.

C. Computational Complexity

The computational complexity orders of the proposed ADVMM and SDVMM algorithms and their comparison with the WAVMAX algorithm [21] are now discussed in this section. For the ADVMM algorithm in Table II, it is easy to verify that each column update involves complexity order $\mathcal{O}((N-1)L)$. Combining all the N columns and denoting the number of alternating cycles required to converge by ζ, the complexity order of the ADVMM is $\mathcal{O}(\zeta N^2 L)$. Regarding the SDVMM algorithm in Table III which only involves simple matrix/vector additions and multiplications, the complexity order can be easily verified as $\mathcal{O}(N^2 L)$. Moreover, as has been analyzed and reported in [21], the complexity order of the WAVMAX algorithm is $\mathcal{O}(N\zeta K((N-1)L + \zeta_w L + \zeta_u N^2 (N-1)^\eta))$, where K is the maximum number of subgradient iterations, ζ and ζ_u are the number of alternating cycles for the outer and inner subproblems, respectively, ζ_w denotes the number of iterations required by the water filling algorithm, and $\eta \in (2.3, 2.8)$. Since the parameters K, ζ, ζ_u, and ζ_w are all positive integers, it is obvious that the complexity of WAVMAX is much higher than that of the proposed ADVMM and SDVMM algorithms. As will be seen in our simulations, the ADVMM and SDVMM not only outperform the WAVMAX in most cases but also would spend much less computation time than the WAVMAX.

VI. COMPUTER SIMULATIONS

In this section, six Monte Carlo simulations are presented to demonstrate the advantages of the proposed RASF algorithm and the ADVMM and SDVMM algorithms.[2] One hundred independent runs were performed in each Monte Carlo simulation. Section VI-A presents the results of the sensitivity of the RASF algorithm to the preassigned number of outliers Z. Section VI-B presents the results of the sensitivity of the ADVMM and SDVMM algorithms to the preassigned back-off tolerance r. In the subsequent sections, we compare the proposed ADVMM and SDVMM algorithms with some existing benchmark nonrobust EEAs, including sequential N-FINDR (SQ-N-FINDR) [36], successive N-FINDR (SC-N-FINDR) [36], simplex growing algorithm (SGA) [35], vertex component analysis (VCA) [38], and the existing robust EEA, WAVMAX [21]. Note that, throughout Sections VI and VII, all the EEAs employed the ASF [17] for dimension reduction unless particularly specified. Section VI-C–E show the performance of the EEAs for various signal-to-noise ratios (SNRs), for various numbers of endmembers, and for various numbers of pixels, respectively. Section VI-F shows the performance of the EEAs with RASF/ASF used over various signal-to-outlier ratios (SORs) (defined in (53) hereinafter). Section VI-G demonstrates the efficacy of the proposed ADVMM/SDVMM with RASF when data contain rare endmembers.

In the simulations, three performance indices were used. The distance between the true affine set $\mathcal{A}(\mathbf{C}, \mathbf{d})$ and the estimated affine set $\mathcal{A}(\hat{\mathbf{C}}, \hat{\mathbf{d}})$, denoted by D_{aff}, for the evaluation of the accuracy of the RASF is defined as

$$D_{\text{aff}} = \frac{\|\mathbf{C}\mathbf{C}^T - \hat{\mathbf{C}}\hat{\mathbf{C}}^T\|_F}{\sqrt{2(N-1)}} + \frac{\left\|\mathbf{P}_{\mathbf{C}}^\perp \mathbf{d} - \mathbf{P}_{\hat{\mathbf{C}}}^\perp \hat{\mathbf{d}}\right\|}{\|\mathbf{P}_{\mathbf{C}}^\perp \mathbf{d}\| + \left\|\mathbf{P}_{\hat{\mathbf{C}}}^\perp \hat{\mathbf{d}}\right\|} \quad (50)$$

where the first term, in the range [0,1], is called the *projection F-norm* [39] and measures the distance between the range space of $\mathbf{C}$ and that of $\hat{\mathbf{C}}$ and the second term in the range [0,1] quantifies the error between $\mathbf{P}_{\mathbf{C}}^\perp \mathbf{d}$ and $\mathbf{P}_{\hat{\mathbf{C}}}^\perp \hat{\mathbf{d}}$. The root mean square (rms) spectral angle distance between the true endmembers and estimated endmembers, denoted by ϕ (in degrees), was used as an accuracy measure of EEAs [38], which is defined as follows:

$$\phi = \min_{\boldsymbol{\pi} \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\arccos\left(\frac{\mathbf{a}_i^T \hat{\mathbf{a}}_{\pi_i}}{\|\mathbf{a}_i\| \cdot \|\hat{\mathbf{a}}_{\pi_i}\|} \right) \right]^2} \quad (51)$$

where $\hat{\mathbf{a}}_i$ denotes the ith estimated endmember signature, $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_N]^T$, and $\Pi_N = \{\boldsymbol{\pi} \in \mathbb{R}^N | \pi_i \in \{1, 2, \ldots, N\}, \pi_i \neq \pi_j \text{ for } i \neq j\}$ is the set of all the permutations of $\{1, 2, \ldots, N\}$. The estimation accuracy defined in (51) with $N!$ permutations $\boldsymbol{\pi}$ can be efficiently solved by the Hungarian algorithm [40]. The smaller the values of D_{aff} (or ϕ), the better the accuracy of the affine set estimate (or the endmember estimates). The computation time T_{sec} (in seconds) of each algorithm (implemented in Mathworks Matlab R2008a) running in a desktop computer equipped with Core i7-930 2.80-GHz CPU and 12-GB memory was used as the computational complexity measure.

Hyperspectral data were synthetically generated independently for each run of the simulation. The contamination-free pixel vectors were generated following the signal model (2) where the endmember signatures with $M = 224$ bands were selected from the U.S. Geological Survey (USGS) Library [41] and the corresponding abundance vectors were generated following the Dirichlet distribution $D(\mathbf{s}[n], \boldsymbol{\mu})$ with $\boldsymbol{\mu} = (1/N)\mathbf{1}_N$ which automatically enforces (A1) and (A2) [38]. Moreover, the N pure pixels were randomly added to the data to enforce (A4). The noisy data were generated by adding independent and identically distributed zero-mean white Gaussian noise to the contamination-free data for different SNRs, where $\text{SNR} = \sum_{n=1}^{L} \|\mathbf{x}[n]\|^2 / (\sigma^2 M L)$. In addition, the outliers were also added to the noisy data, where the outlier indices $\ell_1, \ldots, \ell_Z$ were randomly selected from $\{1, \ldots, L\}$, and the associated outliers were generated by

$$\mathbf{z}[\ell_i] = c\boldsymbol{\kappa}_i, \quad i = 1, \ldots, Z \quad (52)$$

where each element of $\boldsymbol{\kappa}_i$ is a zero-mean unit-variance Laplace random variable and c is a scalar adjusted to satisfy SOR specification, where

$$\text{SOR} = \frac{\sum_{n=1}^{L} \|\mathbf{x}[n]\|^2 / L}{\sum_{i=1}^{Z} \|\mathbf{z}[\ell_i]\|^2 / Z}. \quad (53)$$

The generation of outliers using the Laplace distribution is to fulfill the belief that the outliers should be heavily tailed in distribution, which is highly peaked at zero and falls off more slowly than the Gaussian distribution in the tail. Let us
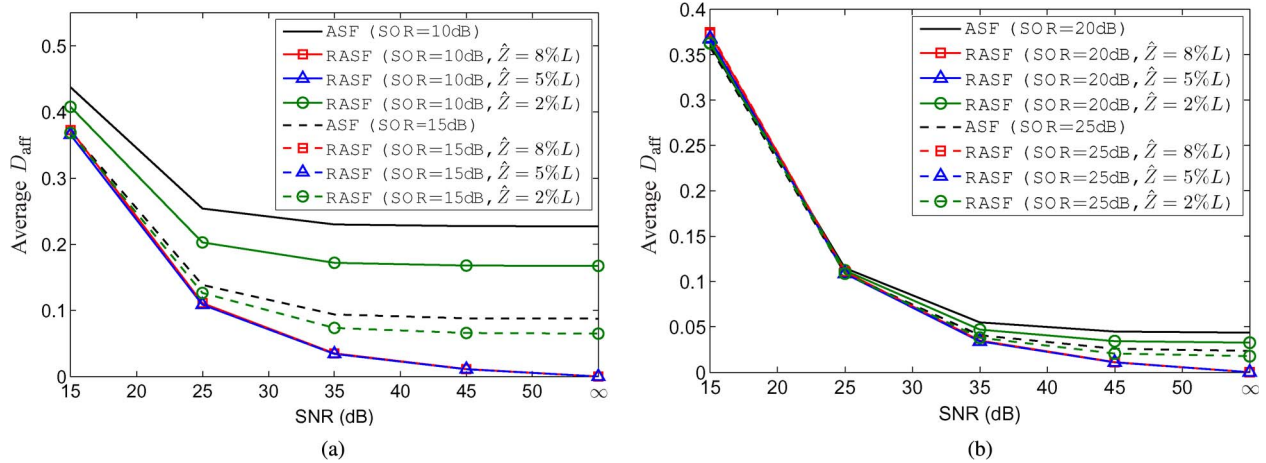
Fig. 3. Performance comparison of ASF and RASF in terms of average D_{aff}, with different preset $\hat{Z}$'s for various SNRs, and $Z = 5\%L$ as well as (a) SOR = 10, 15 dB and (b) SOR = 20, 25 dB.

emphasize again that the proposed RASF algorithm does not require any statistical priors of the outliers. When SNR $\geq$ SOR, the outlier pixels $\mathbf{y}[\ell_1], \ldots, \mathbf{y}[\ell_Z]$ are corrupted by the outliers $\mathbf{z}[\ell_1], \ldots, \mathbf{z}[\ell_Z]$ more seriously than the noise; otherwise, the case of SOR $\geq$ SNR means that the effects of the outliers $\mathbf{z}[\ell_1], \ldots, \mathbf{z}[\ell_Z]$ are smaller than the noise effects, thereby making the outlier pixels $\mathbf{y}[\ell_1], \ldots, \mathbf{y}[\ell_Z]$ not much different from the rest of the observed pixel vectors.

A. RASF for Various SNRs and SORs

In this section, the performance of the proposed RASF and its sensitivity to the preset number of outliers $\hat{Z}$ are evaluated. The hyperspectral data with $N = 8$ number of endmembers (Carnallite, Biotite, Actinolite, Andradite, Clintonite, Diaspore, Goethite, and Halloysite), $L = 1000$ pixels, and $Z = 5\%L$ outliers were generated for SNR = 15, 25, 35, 45, ∞ dB and SOR = 10, 15, 20, 25 dB. The true affine set parameters $(\mathbf{C}, \mathbf{d})$ were obtained from the contamination-free observed pixel vectors $\{\mathbf{x}[n]\}_{n=1}^L$.

Fig. 3 shows the average D_{aff} of the ASF [17] and the proposed RASF algorithm with $\hat{Z} = 2\%L$, $5\%L$, and $8\%L$ for various SORs and SNRs. It can be observed from Fig. 3 that the performance of the RASF algorithm improves as $\hat{Z}$ increases and gets saturated when the preset number of outliers $\hat{Z}$ is greater than the true number of outliers Z. The RASF algorithm also perfectly identifies the true affine set when the noise is absent (i.e., SNR = ∞) and $\hat{Z} \geq Z$. Moreover, the RASF algorithm outperforms ASF for all the values of $\hat{Z}$ under test when SOR $\leq$ SNR, no matter whether SOR is high or low. This implies that, as long as the outlier pixels were corrupted by the outliers more heavily than by the noise, the RASF algorithm will take effect in outlier detection.

B. ADVMM and SDVMM Algorithms for Various Back-Off Tolerances and SNRs

The sensitivity of the proposed ADVMM and SDVMM algorithms to the preset back-off tolerances r is presented. Data generation ($N = 8$, $M = 224$, and $L = 1000$) is the same as that in Section VI-A, where SNR = 15, 25, 35, 45 dB and

SOR = ∞. We set $r = \lambda\sigma$, where σ is the noise standard deviation assumed to be known and λ is a real number varying from 0 to 4 in steps of 0.2.

The average values of ϕ of the ADVMM and SDVMM algorithms versus λ for different SNRs are shown in Fig. 4(a) and (b), respectively. One can see from Fig. 4(a) that the performance of ADVMM is less sensitive to λ over $0 \leq \lambda \leq 3$ for all the SNRs under test, but its performance decreases significantly as $\lambda > 3$ and SNR = 15 dB. It can also be seen from Fig. 4(b) that the sensitivity of SDVMM to λ is low for all the SNRs and λ. Obviously, the sensitivity of SDVMM to both λ and SNR is less than that of ADVMM.

For the performance comparison in Section VI-C–F, we set the back-off tolerance of the proposed ADVMM and SDVMM algorithms to $r = 1.3\sigma$, which could be considered as a modest choice suggested by the simulation results shown in Fig. 4. We also set the same value of r for the robust WAVMAX algorithm [21].

C. EEAs for Various SNRs

Hyperspectral data generation ($N = 8$, $M = 224$, and $L = 1000$) is the same as that in Section VI-A, where we set SNR = 15, 25, 35, 45, ∞ dB and SOR = ∞. Table IV shows the average ϕ and T_{sec} of the EEAs over various SNRs. The minimum ϕ and minimum T_{sec} for a specific SNR over all the algorithms are highlighted as bold-faced numbers. One can see from Table IV that the average ϕ of all the algorithms gradually decreases as the SNR goes up, and they are equal to zero when the SNR approaches to infinity. The performance of SDVMM is the best for almost all the tested SNRs among the existing algorithms under test, and the performances of ADVMM and WAVMAX are quite comparable. On the other hand, the computation time cost by SDVMM is the least, followed by ADVMM, and both are much less than WAVMAX by around an order of 3.

D. EEAs for Various Numbers of Endmembers

The synthetic data generation ($M = 224$ and $L = 1000$) is the same as that in Section VI-A, where SNR = 15 dB, SOR = ∞ dB, and $N = 4, 6, \ldots, 14$ endmembers are randomly selected from the USGS Library [41]. The average ϕ and T_{sec}
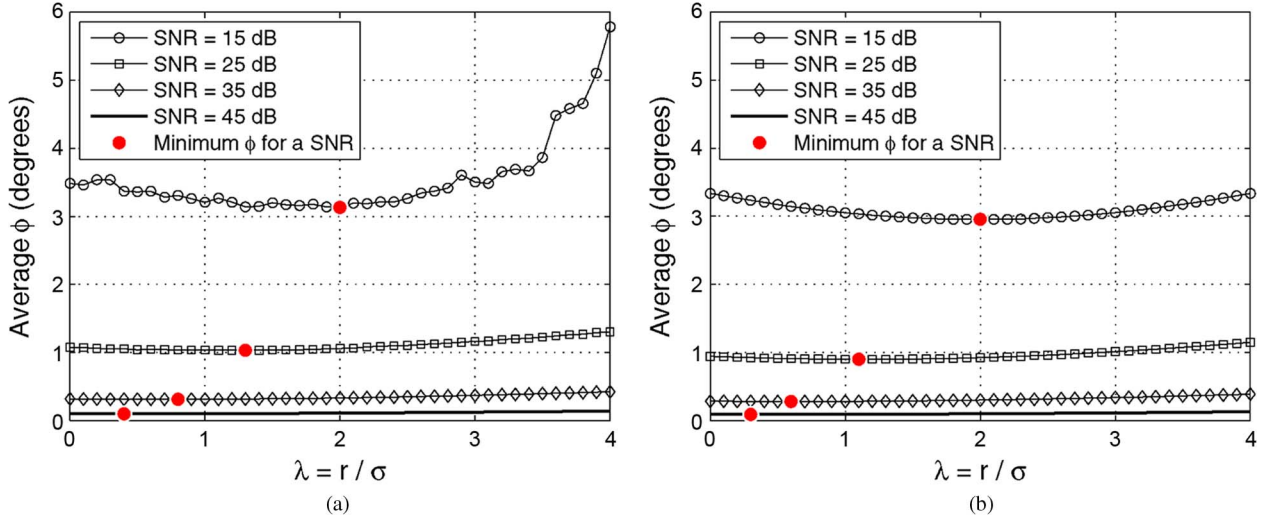
Fig. 4. Performance sensitivity of (a) ADVMM algorithm and (b) SDVMM algorithm w.r.t. λ for different SNRs.

TABLE IV
PERFORMANCE COMPARISON OF AVERAGE ϕ (DEGREES) AND
AVERAGE T_{sec} (SECONDS) OVER SOME EXISTING EEAs FOR
$N = 8$, $L = 1000$, SOR $= \infty$, AND VARIOUS SNRs

Algorithms		SNR (dB)					
		5	15	25	35	45	∞
VCA	ϕ	15.34	3.79	1.26	0.44	0.13	**0.00**
	T_{sec}	0.073	0.053	0.054	0.032	0.028	0.024
SGA	ϕ	13.92	3.33	0.96	**0.28**	**0.09**	**0.00**
	T_{sec}	0.126	0.132	0.127	0.123	0.121	0.120
SQ-N-FINDR	ϕ	14.13	3.49	1.07	0.31	0.10	**0.00**
	T_{sec}	0.179	0.152	0.132	0.119	0.119	0.118
SC-N-FINDR	ϕ	14.90	3.91	1.08	0.34	0.10	**0.00**
	T_{sec}	0.068	0.065	0.060	0.059	0.059	0.059
WAVMAX	ϕ	13.30	3.15	1.03	0.31	0.10	**0.00**
	T_{sec}	42.879	46.383	49.804	38.155	40.921	41.577
ADVMM	ϕ	**12.95**	3.15	1.03	0.31	0.10	**0.00**
	T_{sec}	0.042	0.029	0.018	0.016	0.016	0.015
SDVMM	ϕ	13.50	**3.00**	**0.89**	**0.28**	**0.09**	**0.00**
	T_{sec}	**0.028**	**0.020**	**0.016**	**0.015**	**0.015**	**0.015**

TABLE V
PERFORMANCE COMPARISON OF AVERAGE ϕ (DEGREES) AND AVERAGE
T_{sec} (SECONDS) OVER SOME EXISTING EEAs FOR $L = 1000$,
SNR $= 15$ dB, SOR $= \infty$, AND VARIOUS NUMBERS OF ENDMEMBERS N

Algorithms		The number of endmembers (N)					
		4	6	8	10	12	14
VCA	ϕ	2.27	2.64	3.40	5.78	9.10	9.72
	T_{sec}	0.097	0.120	0.116	0.461	0.769	0.564
SGA	ϕ	1.97	2.36	3.08	6.43	8.65	9.60
	T_{sec}	0.177	0.217	0.184	0.248	0.353	0.384
SQ-N-FINDR	ϕ	2.11	2.58	3.17	4.89	8.58	9.85
	T_{sec}	0.154	0.195	0.171	0.259	0.394	0.505
SC-N-FINDR	ϕ	2.32	2.87	3.52	6.01	9.41	10.44
	T_{sec}	0.089	0.105	0.088	0.118	0.156	0.167
WAVMAX	ϕ	**1.58**	2.17	2.86	5.77	9.12	9.77
	T_{sec}	24.222	58.961	45.110	87.006	197.176	252.528
ADVMM	ϕ	1.59	2.19	2.87	**4.61**	**8.39**	**9.46**
	T_{sec}	0.047	0.057	0.053	0.089	0.165	0.222
SDVMM	ϕ	1.61	**2.08**	**2.73**	5.33	9.05	9.52
	T_{sec}	**0.047**	**0.055**	**0.046**	**0.064**	**0.081**	**0.076**

of the EEAs for the synthetic data with different N's are shown in Table V. One can see that the performances of all the algorithms degrade as N increases. WAVMAX is the best for $N = 4$. For $N = 6, 8$, the SDVMM outperforms all the other algorithms, and ADVMM does so for $N = 10, 12, 14$. Again, the computation time of the proposed two algorithms is the least and, in particular, is much less than that of WAVMAX by an order of 3.

E. EEAs for Various Numbers of Pixels

Again, the synthetic data generation ($N = 8$ and $M = 224$) is the same as that in Section VI-A, where SNR $= 15$ dB, SOR $= \infty$ dB, and the number of pixels L varies from 250 to 8000. The average ϕ and T_{sec} of the EEAs for the synthetic data with different L's are shown in Table VI. One can see from Table VI that the proposed SDVMM and ADVMM outperform all the other algorithms in all cases. On the other hand, the computation times of all the algorithms, except the VCA algorithm, gradually increase as the L increases. The proposed ADVMM and SDVMM algorithms spend much less computation time than WAVMAX by an order of magnitude between 3 and 4 when $L = 8000$.

TABLE VI
PERFORMANCE COMPARISON OF AVERAGE ϕ (DEGREES) AND AVERAGE
T_{sec} (SECONDS) OVER SOME EXISTING EEAs FOR $N = 8$, SNR $= 15$ dB,
SOR $= \infty$, AND VARIOUS NUMBERS OF PIXELS L

Algorithms		The number of pixels (L)					
		250	500	1000	2000	4000	8000
VCA	ϕ	5.61	4.33	3.80	3.45	3.43	3.66
	T_{sec}	0.229	0.202	0.169	0.213	0.246	0.274
SGA	ϕ	5.32	3.95	3.34	3.01	2.97	2.94
	T_{sec}	0.118	0.192	0.255	0.577	1.054	2.231
SQ-N-FINDR	ϕ	5.23	4.08	3.50	3.19	3.11	3.11
	T_{sec}	0.109	0.179	0.247	0.556	0.907	1.664
SC-N-FINDR	ϕ	5.99	4.66	3.92	3.52	3.54	3.34
	T_{sec}	0.066	0.096	0.120	0.244	0.384	0.649
WAVMAX	ϕ	5.24	4.01	3.15	2.92	2.73	2.58
	T_{sec}	11.518	30.861	65.850	233.859	778.585	2997.140
ADVMM	ϕ	5.16	3.87	3.15	2.85	2.70	2.61
	T_{sec}	0.066	0.073	0.072	0.111	0.166	0.218
SDVMM	ϕ	**5.04**	**3.66**	**3.00**	**2.69**	**2.49**	**2.42**
	T_{sec}	**0.049**	**0.056**	**0.060**	**0.099**	**0.149**	**0.184**

F. EEAs With RASF and With ASF Used for Various SORs

The performance difference between EEAs with RASF and with ASF used is evaluated herein. We also include the state-of-the-art robust endmember estimation method, the RULU

TABLE VII
PERFORMANCE COMPARISON OF AVERAGE ϕ (DEGREES) OVER SOME EXISTING EEAS WITH RASF AND WITH ASF FOR $N = 8$, $L = 1000$, $Z = 5\%L$, SNR = 15 dB, AND VARIOUS SORS

Algorithms		SOR (dB)					
		5	8	11	14	17	20
RULU	-	14.61	14.66	15.56	14.85	13.79	13.43
VCA	RASF	3.36	3.40	3.39	3.38	3.38	3.42
	ASF	17.49	10.96	5.44	3.87	3.53	3.65
SGA	RASF	3.13	3.12	3.13	3.13	3.12	3.17
	ASF	**15.46**	**9.59**	4.94	3.43	3.19	3.14
SQ-N-FINDR	RASF	3.22	3.22	3.22	3.22	3.22	3.23
	ASF	17.14	10.34	4.83	3.54	3.28	3.22
SC-N-FINDR	RASF	3.41	3.31	3.40	3.49	3.48	3.39
	ASF	17.70	11.02	5.56	3.92	3.60	3.60
WAVMAX	RASF	2.95	3.01	3.00	2.89	3.15	2.97
	ASF	16.81	10.25	5.14	3.69	3.34	3.35
ADVMM	RASF	2.93	2.90	2.90	2.88	2.90	2.90
	ASF	16.99	9.95	4.86	3.22	2.97	2.93
SDVMM	RASF	**2.76**	**2.77**	**2.77**	**2.76**	**2.76**	**2.77**
	ASF	16.82	9.74	**4.53**	**3.08**	**2.85**	**2.79**

TABLE VIII
PERFORMANCE COMPARISON OF AVERAGE ϕ (DEGREES) OVER ADVMM-RASF AND SDVMM-RASF WITH $\hat{N} = 8, 10$ FOR $N = 10$, $L = 1000$, $Z = 5\%L$, SNR = ∞, 35 dB, AND VARIOUS SORS

Algorithms		SOR (dB) ($\hat{N} = 10, \hat{Z} = 5\%L$)					SOR (dB) ($\hat{N} = 8, \hat{Z} = 6\%L$)				
		8	11	14	17	20	8	11	14	17	20
ADVMM-RASF	ϕ_B	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
(SNR= ∞)	ϕ_R	0.00	0.00	0.00	0.00	0.00	–	–	–	–	–
SDVMM-RASF	ϕ_B	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
(SNR= ∞)	ϕ_R	0.00	0.00	0.00	0.00	0.00	–	–	–	–	–
ADVMM-RASF	ϕ_B	0.31	0.30	0.30	0.31	0.30	0.32	0.32	0.32	0.32	0.32
(SNR = 35 dB)	ϕ_R	3.08	3.08	3.09	3.09	3.08	–	–	–	–	–
SDVMM-RASF	ϕ_B	0.23	0.23	0.23	0.23	0.23	0.30	0.30	0.30	0.30	0.30
(SNR= 35 dB)	ϕ_R	3.13	3.13	3.13	3.13	3.13	–	–	–	–	–

algorithm [16], for our performance comparison. The data generation ($M = 224$, $N = 8$, and $L = 1000$) is the same as that in Section VI-A, where $Z = 5\%L$, SNR is fixed to 15 dB, and SOR = $5, 8, \ldots, 20$ dB. Table VII shows the average ϕ of the EEAs with RASF/ASF used and the RULU algorithm over various SORs. One can see from Table VII that the performances of all the EEAs with ASF used improve as the SOR increases, and the RASF algorithm substantially boosts the performances of all the EEAs in the presence of outliers. Although the performance of EEAs with ASF is worse than that of RULU for low SORs (namely, SOR = 5 dB), EEAs with RASF outperform RULU for all the SORs tested.

Moreover, the performances of any EEAs with RASF for all SORs are quite competitive, meaning that the RASF can be in conjunction with any EEAs to provide better endmember estimates than the ASF. Last but not least, the performance of the SDVMM with RASF used is still the best.

G. ADVMM-RASF and SDVMM-RASF Algorithms for Rare Endmembers

So far, we have demonstrated the superior performance of the RASF algorithm, ADVMM/SDVMM algorithm, and their combinations over some benchmark methods in the previous sections. Now, one may question whether the proposed algorithms can preserve rare endmembers if the data really contain rare endmembers. Following the similar data generation ($M = 224$, $L = 1000$, and $Z = 5\%L$) as in Section VI-A, in addition to the eight endmembers considered as background endmembers, we also add two endmembers, Montmorillonite and Muscovite, as rare endmembers. Each rare endmember occupies only 2 pixels. We then consider two scenarios. First, the number of endmembers is perfectly estimated, i.e., $\hat{N} = 10$, and the given number of outliers $\hat{Z} = Z = 5\%L$. Second, the number of endmembers is underestimated, and it corresponds to the number of background endmembers, for example, $\hat{N} = 8$. Here, the given number of outliers is set to $\hat{Z} = 6\%L$, and making such $1\%L$ increment in $\hat{Z}$ is to account for errors made by the $N - \hat{N}$ endmembers, as only $\hat{N} = 8$ endmembers will be finally extracted. The aforementioned two cases for different $\hat{N}$'s are to simulate various outcomes of methods for estimating

the number of endmembers, such as hyperspectral subspace identification by minimum error (HySime) [23], maximum orthogonal-complement algorithm (MOCA) [30], robust signal subspace estimation [31], modified MOCA [32], and geometry-based estimation of number of endmembers [33].

The performances of ADVMM-RASF and SDVMM-RASF with the generated data for SNR = ∞, 35 dB and various SORs are shown in Table VIII, where ϕ_B (or ϕ_R) is the rms spectral angle between the estimated background (or rare) endmembers and their ground truth. One can easily see from Table VIII that, when $\hat{N} = N = 10$, the proposed algorithms can estimate both background endmembers and rare endmembers with high accuracy in noise-free and noisy cases. When $\hat{N} = 8$, all the estimated endmembers are found to be the background endmembers. Moreover, due to $\hat{N} \neq N$, the proposed algorithms treat rare endmembers as outliers, and their performance slightly degrades. The aforementioned observations suggest that, if one wants to preserve the rare endmembers after dimension reduction, the given $\hat{N}$ should include both the number of background endmembers and the number of rare endmembers. This immediately confirms the importance of the estimation of the number of endmembers in the presence of rare endmembers [30]–[33].

VII. HYPERSPECTRAL DATA EXPERIMENTS

In this experiment, SGA [35], WAVMAX [21], and the proposed ADVMM and SDVMM algorithms, all with ASF used, were tested on the AVIRIS hyperspectral data taken over the Cuprite mining site, NV, in 1997 [42]. The SDVMM with RASF used (SDVMM-RASF) was also applied to the AVIRIS Cuprite data set for comparison. We consider a subimage (200×200 pixels, $L = 40\,000$) of the hyperspectral data as the region of interest which comprises 224 spectral bands over a wavelength from 0.4 to 2.5 μm. The bands 1–2, 104–113, 148–167, and 221–224, which are in low SNR (due to the effect of water vapor), were removed from the original 224-band hyperspectral data. A total of 188 bands were therefore used in our experiment. Moreover, since the noise in real data may not be isotropically distributed, we applied noise prewhitening to the data set, i.e.,

$$\mathbf{y}_{\mathrm{p}}[n] \triangleq \mathbf{D}^{-1/2}\mathbf{y}[n] = \sum_{i=1}^{N} s_i[n]\mathbf{g}_i + \mathbf{z}_{\mathrm{p}}[n] + \mathbf{w}_{\mathrm{p}}[n], \forall n \quad (54)$$

where the noise covariance matrix $\mathbf{D}$ was estimated by the multiple linear regression method [23], $\mathbf{g}_i = \mathbf{D}^{-1/2}\mathbf{a}_i$, $\mathbf{z}_{\mathrm{p}}[n] = \mathbf{D}^{-1/2}\mathbf{z}[n]$, and $\mathbf{w}_{\mathrm{p}}[n] = \mathbf{D}^{-1/2}\mathbf{w}[n]$ is the isotropically distributed noise with the covariance matrix equal to $\mathbf{I}_M$. Once
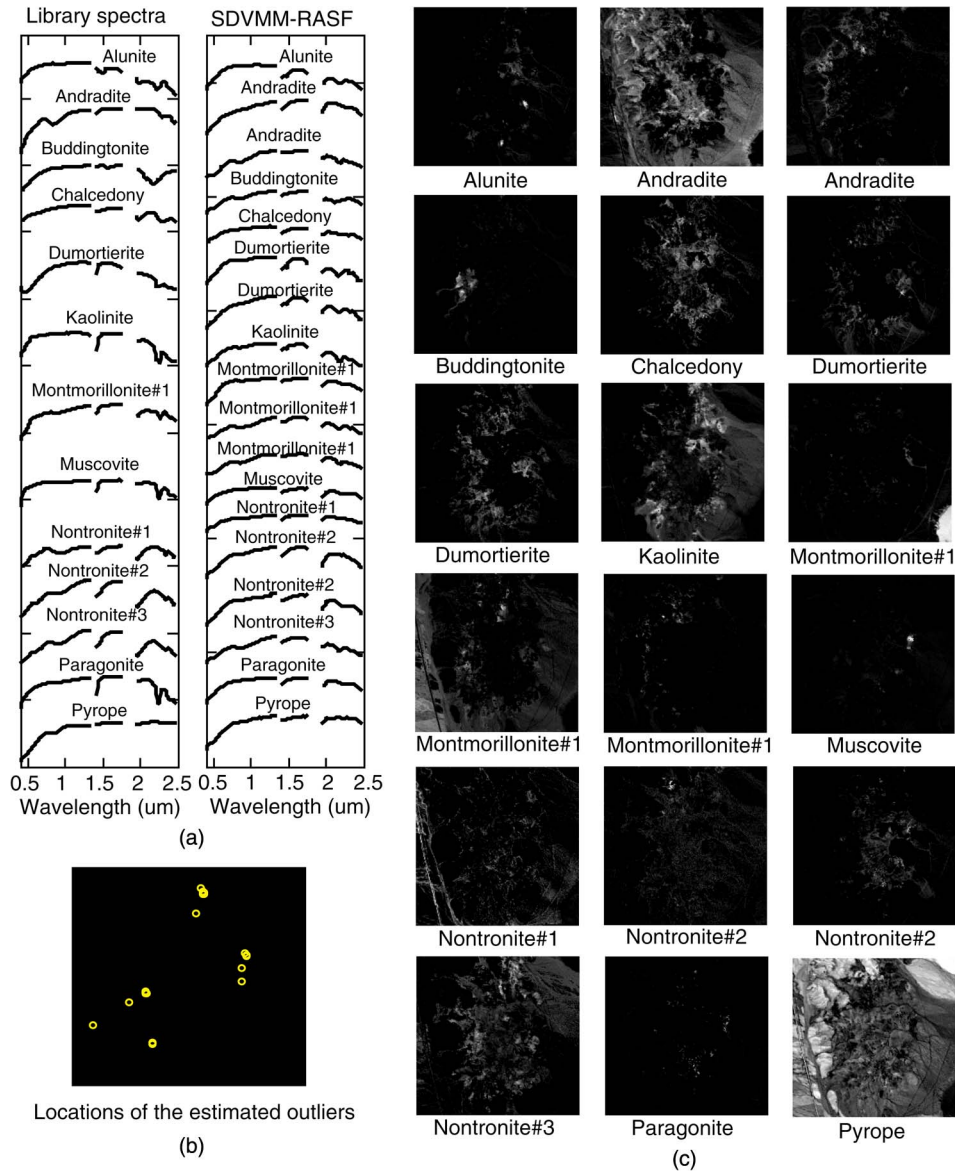
Fig. 5. Experimental results of SDVMM-RASF with Cuprite data set: (a) Endmember signatures taken from the USGS Library, and the estimated endmembers; (b) locations of the $\hat{Z} = 20$ outliers detected by RASF; and (c) 18 abundance maps associated with the estimated endmembers.

$\hat{\mathbf{g}}_1, \ldots, \hat{\mathbf{g}}_N$ were found by any EEA, the endmember estimates can be recovered by $\hat{\mathbf{a}}_i = \mathbf{D}^{1/2}\hat{\mathbf{g}}_i$, $i = 1, \ldots, N$. According to the noise prewhitening processing where the resulting noise variance becomes unity, the back-off tolerance r of the robust WAVMAX, ADVMM, SDVMM, and SDVMM-RASF algorithms was set to 1.3. The HySime [23] was applied to estimate the number of endmembers in this region, and the results yield $N = 18$. The number of outliers $\hat{Z}$ used in RASF was empirically set to $0.05\%L = 20$. How to optimally tune the value of $\hat{Z}$ will be considered as our future directions, but as suggested by simulations, there will be not much performance improvement when $\hat{Z}$ is larger than the (unknown) true Z.

The abundance maps associated with the estimated endmembers for all the EEAs under test were obtained by fully constrained least square method [43]. The minerals were then identified by visual comparison of the obtained abundance maps with the ones in [21], [14], and [38]. Due to the space limit, we herein only demonstrate the endmember estimates

[see Fig. 5(a)], the locations of the $0.05\%L = 20$ detected outliers [see Fig. 5(b)], and the estimated abundance maps [see Fig. 5(c)] for the proposed SDVMM-RASF. The true endmember signatures taken from the USGS Library were also shown in Fig. 5(a) for comparison [44]. In addition, the mean-removed spectral angle (degrees) between the estimated endmember $\hat{\mathbf{a}}$ and the associated USGS Library endmember signature $\mathbf{a}$ was used as the performance measure

$$\bar{\phi} = \arccos\left(\frac{(\hat{\mathbf{a}} - \mathbf{m}(\hat{\mathbf{a}}))^T (\mathbf{a} - \mathbf{m}(\mathbf{a}))}{\|\hat{\mathbf{a}} - \mathbf{m}(\hat{\mathbf{a}})\| \cdot \|\mathbf{a} - \mathbf{m}(\mathbf{a})\|}\right) \quad (55)$$

where $\mathbf{m}(\mathbf{a}) = (\mathbf{1}_M^T \mathbf{a}/M)\mathbf{1}_M$ for any vector $\mathbf{a} \in \mathbb{R}^M$. The smaller the value of $\bar{\phi}$, the better the accuracy of endmember signature estimates. The values of $\bar{\phi}$ for the estimated endmembers yielded by all the various EEAs are shown in Table IX, where the least $\bar{\phi}$ for a specific mineral over those five algorithms is highlighted as a bold-faced number and the

TABLE IX
MEAN-REMOVED SPECTRAL ANGLES $\bar{\phi}$ (DEGREES) BETWEEN LIBRARY SPECTRA AND ENDMEMBERS ESTIMATED BY THE SGA, WAVMAX, ADVMM, AND SDVMM

	SGA	WAVMAX	ADVMM	SDVMM	SDVMM-RASF
Alunite	**18.02**	22.00	21.90	21.48	20.47
Andradite	25.49	22.22	-	21.53	**16.73** (26.43)
Buddingtonite	**24.49**	26.19	24.54	24.53	26.40
Chalcedony	25.98	-	-	**24.00**	**24.00**
Desert Varnish	-	11.79	12.04	**11.64**	-
Dumortierite	24.96 (32.02) (31.54) (21.01)	28.77 (25.42)	25.64 (25.19)	33.53 (27.10)	**20.47** (27.10)
Kaolinite	**24.11**	27.56	27.09	24.14	27.58
Montmorillonite#1	17.14	19.65 (24.58)	18.90	17.53	**17.12** (22.65) (22.36)
Montmorillonite#2	25.70	25.70	25.67	**20.65** (25.70)	-
Muscovite	**35.28**	37.86	37.67	**35.26**	35.61
Nontronite#1	25.04 (**24.84**)	25.69	25.06 (25.77) (25.64)	25.03	25.99
Nontronite#2	16.18	17.00 (23.04)	23.03 (**15.69**)	17.53 (23.01)	17.55 (22.65)
Nontronite#3	25.78	26.96 (26.62)	26.33 (27.70)	25.72	**24.59**
Paragonite	35.36	31.37	**30.92**	35.42	35.36
Pyrope	**13.03**	13.09	13.20	15.73	15.70
Average ϕ	24.78	24.22	24.00	23.93	**23.82**

numbers in parentheses denote the value of $\bar{\phi}$ for the repeatedly estimated endmember. It can be observed from Table IX that the $\bar{\phi}$ of the SDVMM-RASF method is the least, although there is no much difference in $\bar{\phi}$ for all the methods under test. The possible reason is that the quality of the AVIRIS data set is high, i.e., both the SNR and SOR of the AVIRIS Cuprite data set are large [45] and the true outlier pixels are very few. Hence, such $\hat{Z} = 20$ estimated outliers may presumably be the pixels with the $\hat{Z}$ largest approximation errors between the linear spectral mixing model (1) and the nonlinear mixing model of the real hyperspectral data [2], [27]. Moreover, we have also tuned $\hat{Z}$ equal to $0.1\%L$ and $0.3\%L$ for the SDVMM-RASF, and the associated performances do not vary much, inferring that $\hat{Z} = 0.05\%L$ outliers should sufficiently cover all the true outliers, i.e., $\hat{Z}$ is larger than the true Z.

As presented earlier that the AVIRIS Cuprite data set is of very high quality, one may raise a question: Because the information of the outliers is not publicly available, do the locations of the $\hat{Z} = 20$ outliers estimated by the RASF algorithm [see Fig. 5(b)] really cover the locations of the true outliers? To properly investigate this, we generated a semireal hyperspectral data set with outliers artificially added. We selected three spatial coordinates, for example, (16, 24), (24, 16), and (24, 32), from the 200×200 cropped data and artificially added three outliers $\mathbf{z}[n]$, as well as adjusted these outliers to satisfy different SORs defined in (53), provided that the original hyperspectral data were contamination free. Fig. 6(a) displays the locations of the three simulated outliers, marked in yellow triangles on the 50th band hyperspectral image, and one can see that they geometrically and spatially form a triangle. After applying the noise prewhitening (54) to the outlier-added data, we then tested the proposed SDVMM-RASF algorithm with $\hat{Z} = 0.05\%L$ on the data sets of various SORs and show the estimated locations of the outliers in Fig. 6(b) and (c) for SOR ≤ 45 dB and SOR > 45 dB, respectively. For SOR ≤ 45 dB, it can be observed that there is a clear triangle shape formed by the three outlier locations in the upper left corner of Fig. 6(b), showing a good consistency with our simulated outlier locations in Fig. 6(a). For the scenario where SOR > 45 dB, we cannot identify any preplaced outlier locations in Fig. 6(c), conceivably because either the SNR of the Cuprite data set is around 45 dB or the
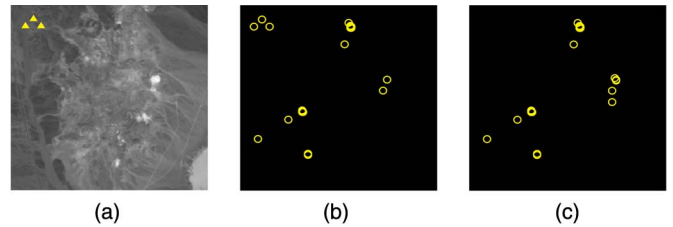

Fig. 6. (a) Locations of the generated outliers marked on the 50th band hyperspectral image and the locations of the estimated outliers by the RASF algorithm for (b) SOR ≤ 45 dB and (c) SOR > 45 dB.

outliers (marked by circles) in the hyperspectral data are more dominant (i.e., the corresponding SOR is lower than 45 dB) than the three preplaced outliers. Besides, the performance of the estimated endmembers by the SDVMM-RASF algorithm is the same as that in Table IX, and hence, the associated results are not demonstrated herein.

VIII. CONCLUSION

We have presented a two-step approach for robust hyperspectral endmember extraction in the presence of outliers and noise. For dimension reduction, the RASF algorithm was proposed to find a robust affine set and, meanwhile, to detect and remove the outliers. For endmember extraction, we have also presented two fast algorithms, namely, ADVMM and SDVMM, to approximate the simplex volume max-min formulation that has shown its robustness against noise [21]. The proposed ADVMM and SDVMM algorithms decouple the max-min simplex volume problem and solve the partial max-min problems in alternating fashion and successive fashion, respectively. The decoupled max-min heuristic and approximation mechanisms used in the proposed ADVMM and SDVMM algorithms have also been discussed. All the subproblems involved in the RASF algorithm, as well as the ADVMM and SDVMM algorithms, end up with closed-form solutions, and hence, the proposed methods are computationally efficient.

Monte Carlo simulation results have shown that the RASF algorithm provides a more accurate affine set estimate than the ASF, as long as SOR $\leq$ SNR. The simulation results have

also shown that the ADVMM and SDVMM algorithms perform slightly better than the robust algorithm WAVMAX [21] while outperforming some existing benchmark algorithms, particularly for low SNR. The computation load in terms of running time required by the ADVMM and SDVMM algorithms is significantly less than that of the WAVMAX algorithm by around an order of 3. Moreover, the results also demonstrated that the performances of any EEAs preceded by the dimension reduction using RASF improve significantly, particularly for low SOR. In cases where some of the endmembers are rare, the proposed ADVMM/SDVMM-RASF algorithms can still preserve the rare endmembers if the number of endmembers is perfectly given, i.e., $\hat{N} = N$. For $\hat{N} < N$, the rare endmembers may instead be treated as outliers. Real data experiments using the Cuprite data set have also demonstrated the practical applicability of the proposed RASF, ADVMM, and SDVMM algorithms.

APPENDIX
PROOF OF LEMMA 1

Problem (44) is equivalent to

$$\min_{\|\mathbf{t}_j\|_2 \le r} \left\| \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} (\mathbf{w}_j - \mathbf{t}_j) \right\|^2 \tag{56}$$

whose KKT conditions can be easily shown to be

$$\left(\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} + \hat{\lambda} \mathbf{I}_N \right) \hat{\mathbf{t}}_j = \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j, \tag{57a}$$

$$\hat{\lambda} \left(\|\hat{\mathbf{t}}_j\|^2 - r^2 \right) = 0, \tag{57b}$$

$$\|\hat{\mathbf{t}}_j\|^2 - r^2 \le 0, \quad \hat{\lambda} \ge 0, \tag{57c}$$

where $\hat{\mathbf{t}}_j$ and $\hat{\lambda}$ are primal and dual optimal points of (56). By (57a) and (57c), we have

$$\left\| \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j \right\| \le \left\| \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} + \hat{\lambda} \mathbf{I}_N \right\| \cdot \|\hat{\mathbf{t}}_j\| \le (1+\hat{\lambda}) r \tag{58}$$

where the first inequality is due to the inequality of the operator norm and the second inequality is due to the fact that the eigenvalues of a projection matrix are equal to either zero or one.

To prove Lemma 1, we first prove

$$\hat{\lambda} > 0 \iff \left\| \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j \right\| > r. \tag{59}$$

First, let us show the sufficiency of (59). Suppose that $\hat{\lambda} > 0$. Then, $\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} + \hat{\lambda} \mathbf{I}_N$ is of full rank, and (57a) becomes

$$\hat{\mathbf{t}}_j = \left(\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} + \hat{\lambda} \mathbf{I}_N \right)^{-1} \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j. \tag{60}$$

By eigenvalue decomposition, we can express

$$\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} = \mathbf{V} \begin{bmatrix} \mathbf{I}_{N-(j-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^T \tag{61}$$

where $\mathbf{V} \in \mathbb{R}^{N \times N}$ is the matrix constituted by orthonormal eigenvectors of $\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}}$. Substituting (61) into (60) followed by some derivations, (60) can be simplified to

$$\hat{\mathbf{t}}_j = \frac{1}{1 + \hat{\lambda}} \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j. \tag{62}$$

By (62) and (57b), it is easy to see $\|\hat{\mathbf{t}}_j\| = \|(1/(1+\hat{\lambda})) \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j\| = r$, which yields

$$\hat{\lambda} = \frac{1}{r} \left\| \mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j \right\| - 1 \tag{63}$$

thus leading to $\|\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j\| > r$ due to $\hat{\lambda} > 0$. The proof of necessity of (59) is trivial. By (58), $\|\mathbf{P}^{\perp}_{\widehat{\mathbf{H}}_{1:(j-1)}} \mathbf{w}_j\| > r$ implies $\hat{\lambda} > 0$.

Now, we make use of (59) to prove Lemma 1. Two cases on $\mathbf{w}_j \in \mathcal{F}$ are considered: (C1) $\mathbf{w}_j \in \mathcal{W}(r)$ and (C2) $\mathbf{w}_j \in \mathbb{R}^N \setminus \mathcal{W}(r)$. By (59), (C1) implies $\hat{\lambda} > 0$. Hence, by (62) and (63), the solution (45) can be obtained. On the other hand, for case (C2), by (57c) and (59), (C2) implies $\hat{\lambda} = 0$. Hence, (46) can be obtained by setting $\hat{\lambda} = 0$ in (57a). ∎

REFERENCES

[1] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ: Wiley, 2007.
[2] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
[3] 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, Jun. 2012.
[4] J. M. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing: Geometrical, statistical, and sparse regression-based approaches," in *Proc. SPIE—Image Signal Process. Remote Sens. XVI*, Toulouse, France, Sep. 20, 2010, vol. 7830, no. 7830A.
[5] A. Plaza, P. Martinez, R. Perez, 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.
[6] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, P. Marcoionni, I. Pippi, and M. Selva, "Noise modelling and estimation of hyperspectral data from airborne imaging spectrometers," *Ann. Geophys.*, vol. 49, no. 1, pp. 1–9, Feb. 2006.
[7] J. Meola, M. T. Eismann, R. L. Moses, and J. N. Ash, "Modeling and estimation of signal-dependent noise in hyperspectral imagery," *Appl. Opt.*, vol. 50, no. 21, pp. 3829–3846, Jul. 2011.
[8] F. A. Kruse, L. L. Richrson, and V. G. Ambrosia, "Techniques for geologic analysis of hyperspectral data applied to near-shore hyperspectral data," in *Proc. 4th Int. Conf. Remote Sens. Marine Coastal Environ.*, Orlando, FL, Mar. 17–19, 1997, vol. 1, pp. 233–246.
[9] T. Han, D. G. Goodenough, A. Dyk, and J. Love, "Detection and correction of abnormal pixels in Hyperion images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Toronto, ON, Canada, Jun. 25–26, 2002, vol. 3, pp. 1327–1330.
[10] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[11] T. E. Smetek and K. W. Bauer, "A comparison of multivariate outlier detection methods for finding hyperspectral anomalies," *Mil. Oper. Res.*, vol. 13, no. 4, pp. 19–43, 2008.

[12] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[13] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st IEEE Workshop Hyperspec. Image Signal Process., Evol. Remote Sens.*, Grenoble, France, Aug. 26–28, 2009, pp. 1–4.

[14] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance constrained robust minimum volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4194–4209, Nov. 2011.

[15] A. M. Filippi and R. Archibald, "Support vector machine-based endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 771–791, Mar. 2009.

[16] O. Duran and M. Petrou, "Robust endmember extraction in the presence of anomalies," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1986–1996, Jun. 2011.

[17] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[18] G. Martin, A. Plaza, and M. Zortea, "Noise-robust spatial preprocessing prior to endmember extraction from hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Vancouver, BC, Canada, Jul. 24–29, 2011, pp. 1287–1290.

[19] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrom.*, Pasadena, CA, Oct. 1999, pp. 266–275.

[20] M. E. Winter, "A proof of the N-FINDR algorithm for the automated detection of endmembers in a hyperspectral image," in *Proc. SPIE Conf. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imag.*, Aug. 2004, vol. 5425, pp. 31–41.

[21] 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.

[22] T.-H. Chan, W.-K. Ma, C.-Y. Chi, and Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5120–5134, Oct. 2008.

[23] 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.

[24] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2957–2970, Aug. 2011.

[25] Q. Du, L. Zhang, and N. Raksuntorn, "Improving the quality of extracted endmembers," in *Proc. 1st IEEE WHISPERS*, Grenoble, France, Aug. 26–28, 2009, pp. 1–4.

[26] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[27] B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[28] M. T. Eismann and R. C. Hardie, "Application of the stochastic mixing model to hyperspectral resolution enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1924–1933, Sep. 2004.

[29] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[30] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.

[31] N. Acito, M. Diani, and G. Corsini, "A new algorithm for robust estimation of the signal subspace in hyperspectral images in presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3844–3856, Nov. 2009.

[32] N. Acito, M. Diani, and G. Corsini, "Hyperspectral signal subspace identification in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940–1954, Apr. 2010.

[33] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Hyperspectral data geometry based estimation of number of endmembers using p-norm based pure pixel identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2753–2769, May 2013.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[35] C.-I. Chang, C.-C. Wu, W.-M. Liu, and Y.-C. Quyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.

[36] C.-C. Wu, S. Chu, and C.-I. Chang, "Sequential N-FINDR algorithms," in *Proc. SPIE*, Aug. 2008, vol. 7086, p. 70860C.

[37] J. V. Neumann, "Zur theorie der gesellschaftsspiele," *Math. Ann.*, vol. 100, no. 1, pp. 295–320, 1928.

[38] 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.

[39] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Apr. 1999.

[40] H. W. Kuhn, "The Hungarian method for the assignment method," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

[41] [Online]. Available: http://speclab.cr.usgs.gov/cuprite.html

[42] AVIRIS Free Standard Data Products. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[43] D. Heinz and C.-I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[44] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. Geological Survey Digital Spectral Library: Version 1: 0.2 to 3.0 μm," U.S. Geol. Surv., Washington, DC, Open File Rep. 93-592, 1993.

[45] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1388–1400, Jun. 2003.
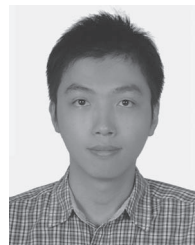
Tsung-Han Chan (S'08–M'09) received the B.S. degree from Yuan Ze University, Taoyuan, Taiwan, in 2004 and the Ph.D. degree from the Institute of Communications Engineering, National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2009.

He is currently a Research Scientist with the Advanced Digital Sciences Center, Singapore. He was a Visiting Doctoral Graduate Research Assistant with Virginia Polytechnic Institute and State University, Arlington, in 2008, and a Postdoctoral Research Fellow with NTHU from 2009 to 2012. His research interests are in signal processing, image processing, and convex optimization, with a recent emphasis on hyperspectral remote sensing, dynamic medical imaging, and computer vision applications.

Dr. Chan is a corecipient of a Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) 2011 Best Paper Award.

ArulMurugan Ambikapathi (S'02–M'11) received the B.E degree in electronics and communication engineering from Bharathidasan University, Tiruchirappalli, India, in 2003, the M.E. degree in communication systems from Anna University, Chennai, India, in 2005, and the Ph.D. degree from the Institute of Communications Engineering (ICE), National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2011. He secured top university ranks in both his B.E. and M.E. programs.

He is currently a Postdoctoral Research Fellow with the ICE, NTHU. His research interests are in convex analysis and optimization for blind source separation, hyperspectral unmixing, and, recently, biomedical image analysis.

Dr. Ambikapathi was the recipient of the Gold medal and Silver medal for academic excellence in his B.E. and M.E. programs, respectively. He is the recipient of the NTHU Outstanding Student Scholarship award for two consecutive years (in 2009 and 2010). He is also a corecipient of a Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing 2011 Best Paper Award.

Wing-Kin Ma (M'01–SM'11) received the B.Eng. (first-class honors) degree in electrical and electronic engineering from the University of Portsmouth, Portsmouth, U.K., in 1995 and the M.Phil. and Ph.D. degrees in electronic engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 1997 and 2001, respectively. His Ph.D. dissertation was commended to be "of very high quality and well deserved honorary mentioning" by the Faculty of Engineering, CUHK, in 2001.

He is currently an Assistant Professor with the Department of Electronic Engineering, CUHK. From 2005 to 2007, he was also an Assistant Professor with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, where he is still holding an adjunct position. Prior to becoming a faculty member, he held various research positions with McMaster University, Hamilton, ON, Canada; CUHK; and the University of Melbourne, Parkville, Vic., Australia. His research interests are in signal processing and communications, with a recent emphasis on multiple-input and multiple-output communication, convex optimization, and blind signal processing.

Dr. Ma currently serves as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. From 2005 to 2009, he has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is also serving or has served as a Guest Editor of the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS on the Special Issue titled "Signal Processing Techniques for Wireless Physical Layer Security," and the IEEE Signal Processing Magazine on the Special Issue titled "Convex Optimization for Signal Processing" in May 2010. He is currently a member of the Signal Processing Theory and Methods Technical Committee. He was a Tutorial Speaker in European Signal Processing Conference (EUSIPCO) 2011 titled "Semidefinite Relaxation of Nonconvex Quadratic Optimization: A Key Technique in Signal Processing and Communications" (cotaught with Anthony M.-C. So). His student has won an International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2011 Best Student Paper Award, and he is a corecipient of a WHISPERS 2011 Best Paper Award.

Chong-Yung Chi (S'83–M'83–SM'89) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1983.

From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA. He has been a Professor with the Department of Electrical Engineering since 1989 and the Institute of Communications Engineering (ICE) since 1999 (also the Chairman of ICE during 2002–2005) at National Tsing Hua University, Hsinchu, Taiwan. He has published more than 190 technical papers, including more than 65 journal papers (mostly in IEEE TRANSACTIONS ON SIGNAL PROCESSING), 2 book chapters, and more than 120 peer-reviewed conference papers, as well as a graduate-level textbook, *Blind Equalization and System Identification*, (Springer-Verlag, 2006). His current research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, and biomedical and hyperspectral image analysis.

Dr. Chi has been a Technical Program Committee member for many IEEE sponsored and cosponsored workshops, symposiums, and conferences on signal processing and wireless communications, including the Co-organizer and General Cochairman of the 2001 IEEE Workshop on Signal Processing Advances in Wireless Communications, as well as the Cochair of the Signal Processing for Communications (SPC) Symposium of International Conference on Communications and Networking in China (ChinaCOM) 2008 and the Lead Cochair of the SPC Symposium of ChinaCOM 2009. He was an Associate Editor (AE) of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (5/2001–4/2006), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (1/2006–12/2007), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I (1/2008–12/2009), and IEEE SIGNAL PROCESSING LETTERS (6/2006–5/2010), a member of the editorial board of the *EURASIP Signal Processing Journal* (6/2005–5/2008), and an Editor (7/2003–12/2005) as well as a Guest Editor (2006) of the *EURASIP Journal on Applied Signal Processing*. He was a member of the IEEE Signal Processing Committee on Signal Processing Theory and Methods (2005–2010). Currently, he is a member of the IEEE Signal Processing Committee on Signal Processing for Communications and Networking, a member of the IEEE Signal Processing Committee on Sensor Array and Multichannel Signal Processing, and an AE of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.