A Fast Hyperplane-Based Minimum-Volume Enclosing Simplex Algorithm for Blind Hyperspectral Unmixing

Chia-Hsiang Lin*, Chong-Yung Chi, Yu-Hsiang Wang, and Tsung-Han Chan

Abstract

Hyperspectral unmixing (HU) is a crucial signal processing procedure to identify the underlying materials (or endmembers) and their corresponding proportions (or abundances) from an observed hyperspectral scene. A well-known blind HU criterion, advocated by Craig in early 1990's, considers the vertices of the minimum-volume enclosing simplex of the data cloud as good endmember estimates, and it has been empirically and theoretically found effective even in the scenario of no pure pixels. However, such kind of algorithms may suffer from heavy simplex volume computations in numerical optimization, etc. In this work, without involving any simplex volume computations, by exploiting a convex geometry fact that a simplest simplex of N vertices can be defined by N associated hyperplanes, we propose a fast blind HU algorithm, for which each of the N hyperplanes associated with the Craig's simplex of N vertices is constructed from $N - 1$ affinely independent data pixels, together with an endmember identifiability analysis for its performance support. Without resorting to numerical optimization, the devised algorithm searches for the $N(N-1)$ active data pixels via simple linear algebraic computations, accounting for its computational efficiency. Monte Carlo simulations and real data experiments are provided to demonstrate its superior efficacy over some benchmark Craig-criterion-based algorithms in both computational efficiency and estimation accuracy.

Index Terms—Hyperspectral unmixing, Craig's criterion, convex geometry, minimum-volume enclosing simplex, hyperplane

I. INTRODUCTION

Hyperspectral remote sensing (HRS) [2]–[4], also known as imaging spectroscopy, is a crucial technology to the identification of material substances (or endmembers) as well as their corresponding fractions (or abundances) present in a scene of interest from observed hyperspectral data, having various applications such as planetary exploration, land mapping and classification, environmental monitoring, and mineral identification and quantification [5]–[7]. The observed pixels in the hyperspectral imaging

data cube are often spectral mixtures of multiple substances, the so-called *mixed pixel* phenomenon [8], owing to the limited spatial resolution of the hyperspectral sensor (usually equipped on board the satellite or aircraft) utilized for recording the electromagnetic scattering patterns of the underlying materials in the observed hyperspectal scene over about several hundreds of narrowly spaced (typically, 5-10 nm) wavelengths that contiguously range from visible to near-infrared bands. Occasionally, the mixed pixel phenomenon can result from the underlying materials intimately mixed [9]. *Hyperspectral unmixing* (HU) [8], [10], an essential procedure of extracting individual spectral signatures of the underlying materials in the captured scene from these measured spectral mixtures, is therefore of paramount importance in the HRS context.

Blind HU, or unsupervised HU, involves two core stages, namely endmember extraction and abundance estimation, without (or with very limited) prior knowledge about the endmembers' nature or the mixing mechanism. Some endmember extraction algorithms (EEAs), such as alternating projected subgradients (APS) [11], joint Bayesian approach (JBA) [12], and iterated constrained endmembers (ICE) [13] (also the sparsity promoting ICE (SPICE) [14]), can simultaneously determine the associated abundance fractions while extracting the endmember signatures. Nevertheless, some EEAs perform endmember estimation, followed by abundance estimation using such as the fully constrained least squares (FCLS) [15] to complete the entire HU processing.

The pure-pixel assumption has been exploited in devising fast blind HU algorithms to search for the purest pixels over the data set as the endmember estimates, and such searching procedure can always be carried out through simple linear algebraic formulations; see, e.g., pixel purity index (PPI) [16] and vertex component analysis (VCA) [17]. An important blind HU criterion, called Winter's criterion [18], also based on the pure-pixel assumption, is to identify the vertices of the maximum-volume simplex inscribed in the observed data cloud as endmember estimates. HU algorithms in this category include N-finder (N-FINDR) [18], simplex growing algorithm (SGA) [19] (also the real-time implemented SGA [20]), and worst case alternating volume maximization (WAVMAX) [21], to name a few. However, the pure-pixel assumption could be seriously infringed in practical scenarios especially when the pixels are intimately mixed, for instance, the hyperspectral imaging data for retinal analysis in the ophthalmology [9]. In these scenarios, HU algorithms in this category could completely fail; actually, it is proven that perfect endmember identifiability is impossible for Winter-criterion-based algorithms if the pure-pixel assumption is violated [21].

Without relying on the existence of pure pixels, another promising blind HU approach, advocated by Craig in early 1990's [22], exploits the simplex structure of hyperspectral data, and believes that

the vertices of the minimum-volume data-enclosing simplex can yield good endmember estimates, and algorithms developed accordingly include such as minimum-volume transform (MVT) [22], minimum-volume constrained nonnegative matrix factorization (MVC-NMF) [23], and minimum-volume-based elimination strategy (MINVEST) [24]. Moreover, some linearization-based methods have also been reported to practically identify Craig's minimum-volume simplex, e.g., the iterative linear approximation in minimum-volume simplex analysis (MVSA) [25] (also its fast implementation using the interior-point method [26], termed as ipMVSA [27]), and the alternating linear programming in minimum-volume enclosing simplex (MVES) [28]. Empirical evidences do well support that this minimum-volume approach is resistant to lack of pure pixels, and can recover ground truth endmembers quite accurately even when the observed pixels are heavily mixed. Very recently, the validity of this empirical belief has been theoretically justified; specifically, we show that, as long as a key measure concerning the pixels' mixing level is above a certain (small) threshold, Craig's simplex can perfectly identify the true endmembers in the noiseless scenario [29]. However, without the guidance of the pure-pixel assumption, this more sophisticated criterion would generally lead to more computationally expensive HU algorithms. To the best of our knowledge, the ipMVSA algorithm [27] and the simplex identification via split augmented Lagrangian (SISAL) algorithm [30] are the two state-of-the-art Craig-criterion-based algorithms in terms of computational efficiency. Nevertheless, in view of not only the NP-hardness of the Craig-simplex-identification (CSI) problem [31] but also heavy simplex volume computations, all the above mentioned HU algorithms are yet to be much more computationally efficient. Moreover, their performances may not be very reliable owing to the sensitivity to regularization parameter tuning, non-deterministic (i.e., non-reproducible) endmember estimates caused by random initializations, and, most seriously, lack of rigorous identifiability analysis.

In this work, we break the deadlock on the trade-off between a simple fast algorithmic scheme and the estimation accuracy in the no pure-pixel case. We have observed that when the pure-pixel assumption holds true, the effectiveness of a simple fast HU algorithmic scheme could be attributed to that the desired solutions (i.e., pure pixels) already exist in the data set. Inspired by this observation, we naturally raise a question: *Can Craig's minimum-volume simplex be identified by simply searching for a specific set of pixels in the data set regardless of the existence of pure pixels?* The answer is affirmative and will be given in this paper.

Based on the convex geometry fact that a simplest simplex of N vertices can be characterized by the N associated hyperplanes, this paper proposes an efficient and effective unsupervised Craig-criterion-based HU algorithm, together with an endmember identifiability analysis. Each hyperplane, parameterized

by a normal vector and an inner product constant [26], can then be estimated from $N - 1$ affinely independent pixels in the data set via simple linear algebraic formulations. The resulting hyperplane-based CSI (HyperCSI) algorithm, based on the above pixel search scheme, can withstand the no pure-pixel scenario, and can yield deterministic, non-negative, and, most importantly, accurate endmember estimates. After endmember estimation, a closed-form expression in terms of the identified hyperplanes' parameters is derived for abundance estimation. Then some Monte Carlo numerical simulations and real hyperspectral data experiments are presented to demonstrate the superior efficacy of the proposed HyperCSI algorithm over some benchmark Craig-criterion-based HU algorithms in both estimation accuracy and computational efficiency.

The remaining part of this paper is organized as follows. In Section II, we briefly review some essential convex geometry concepts, followed by the signal model and dimension reduction. Section III focuses on the HyperCSI algorithm development, and in Section IV, some simulation results are presented for its performance comparison with some benchmark Craig-criterion-based HU algorithms. In Section V, we further evaluate the effectiveness of the proposed HyperCSI algorithm with AVIRIS [32] data experiments. Finally, we draw some conclusions in Section VI.

The following notations will be used in the ensuing presentation. $\mathbb{R}$ ($\mathbb{R}^N$, $\mathbb{R}^{M \times N}$) is the set of real numbers (N-vectors, $M \times N$ matrices). $\mathbb{R}_+$ ($\mathbb{R}_+^N$, $\mathbb{R}_+^{M \times N}$) is the set of non-negative real numbers (N-vectors, $M \times N$ matrices). $\mathbb{R}_{++}$ ($\mathbb{R}_{++}^N$, $\mathbb{R}_{++}^{M \times N}$) is the set of positive real numbers (N-vectors, $M \times N$ matrices). $\mathbf{X}^\dagger$ denotes the Moore-Penrose pseudo-inverse of a matrix $\mathbf{X}$. $\mathbf{1}_N$ and $\mathbf{0}_N$ are all-one and all-zero N-vectors, respectively. $\mathbf{e}_i$ denotes the unit vector of proper dimension with the ith entry equal to unity. $\mathbf{I}_N$ is the $N \times N$ identity matrix. $\succeq$ and $\succ$ stand for the componentwise inequality and strictly componentwise inequality, respectively. $\| \cdot \|$ denotes the Euclidean norm. The distance of a vector $\mathbf{v}$ to a set $\mathcal{S}$ is denoted by $\mathrm{dist}(\mathbf{v}, \mathcal{S}) \triangleq \inf_{\boldsymbol{v} \in \mathcal{S}} \|\mathbf{v} - \boldsymbol{v}\|$ [26]. $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$. The determinant of matrix $\mathbf{X}$ is represented by $\det(\mathbf{X})$. $\mathcal{I}_Z$ stands for the set of integers $\{1, \ldots, Z\}$, for any positive integer Z.

II. Convex Geometry and Signal Model

In this section, a brief review on some essential convex geometry will be given for ease of later use. Then the signal model for representing the hyperspectral imaging data together with dimension reduction preprocessing will be presented.

A. Convex Geometry Preliminary

The *convex hull* of a given set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \subseteq \mathbb{R}^M$ is defined as [26]
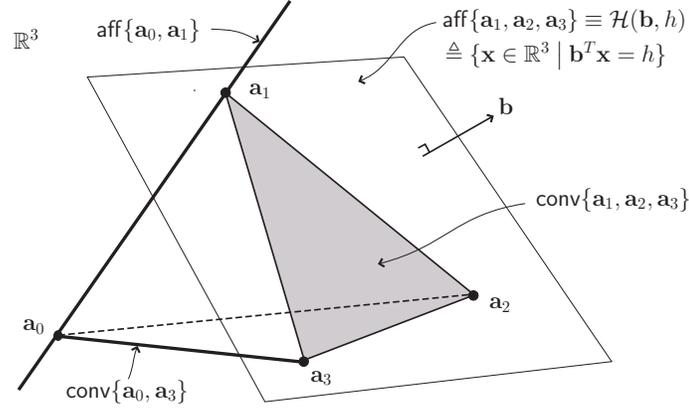
Fig. 1. A graphical illustration in $\mathbb{R}^3$ for some convex geometry concepts, where the line segment connecting $\mathbf{a}_0$ and $\mathbf{a}_3$ is the convex hull of $\{\mathbf{a}_0, \mathbf{a}_3\}$, the straight line passing $\mathbf{a}_0$ and $\mathbf{a}_1$ is the affine hull of $\{\mathbf{a}_0, \mathbf{a}_1\}$, the shaded triangle is the convex hull of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, and the plane passing the three points $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is the affine hull of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. As an affine hull in $\mathbb{R}^3$ is called a hyperplane if its affine dimension is 2, $\mathrm{aff}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is a hyperplane, while $\mathrm{aff}\{\mathbf{a}_0, \mathbf{a}_1\}$ is not.

$$\mathrm{conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \triangleq \left\{ \mathbf{x} = \sum_{i=1}^{N} \theta_i \mathbf{a}_i \ \bigg| \ \boldsymbol{\theta} \in \mathbb{R}_+^N, \mathbf{1}_N^T \boldsymbol{\theta} = 1 \right\},$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T$ (cf. Figure 1). A convex hull $\mathrm{conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ is called an $(N-1)$-dimensional *simplex* with N vertices $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ if $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ is *affinely independent*, or, equivalently, if $\{\mathbf{a}_1 - \mathbf{a}_N, \ldots, \mathbf{a}_{N-1} - \mathbf{a}_N\}$ is linearly independent, and it is called a *simplest simplex* in $\mathbb{R}^M$ when $M = N - 1$ [33]. For example, a triangle is a 2-dimensional simplest simplex in $\mathbb{R}^2$, and a tetrahedron is a 3-dimensional simplest simplex in $\mathbb{R}^3$ (cf. Figure 1).

For a given set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \subseteq \mathbb{R}^M$, its *affine hull* is defined as [26]

$$\mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \triangleq \left\{ \mathbf{x} = \sum_{i=1}^{N} \theta_i \mathbf{a}_i \ \bigg| \ \boldsymbol{\theta} \in \mathbb{R}^N, \mathbf{1}_N^T \boldsymbol{\theta} = 1 \right\},$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T$ (cf. Figure 1). This affine hull can be parameterized by a 2-tuple $(\mathbf{C}, \mathbf{d}) \in \mathbb{R}^{M \times P} \times \mathbb{R}^M$ using the following alternative representation [26]:

$$\mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \equiv \mathcal{A}(\mathbf{C}, \mathbf{d}) \triangleq \left\{ \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \ \big| \ \boldsymbol{\alpha} \in \mathbb{R}^P \right\},$$

where $P \triangleq \mathrm{rank}(\mathbf{C})$ (the rank of $\mathbf{C}$) is the *affine dimension* of $\mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$. Moreover, an affine hull $\mathrm{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \subseteq \mathbb{R}^M$ is called a *hyperplane* if its affine dimension $P = M - 1$ (cf. Figure 1).

B. Signal Model and Dimension Reduction

Consider a scenario where a hyperspectral sensor measures solar electromagnetic radiation over M spectral bands from N unknown materials (endmembers) in a scene of interest. Based on the linear mixing

model (LMM) [2]–[8], [10], [28], where the measured solar radiations are assumed to reflect from the explored scene through one single bounce, and the endmembers' spectral signature vectors $\mathbf{a}_i \in \mathbb{R}^M$ are assumed to be invariant with the pixel index n, each pixel $\mathbf{x}[n] \in \mathbb{R}^M$ in the observed data set can then be represented as a linear mixture of the N endmembers' spectral signatures [1]

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] = \sum_{i=1}^{N} s_i[n]\mathbf{a}_i, \quad \forall n \in \mathcal{I}_L, \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_N] \in \mathbb{R}^{M \times N}$ is the spectral signature matrix, $\mathbf{s}[n] = [s_1[n] \cdots s_N[n]]^T \in \mathbb{R}^N$ is the abundance vector, and L is the total number of pixels. The following standard assumptions pertaining to the model in (1), which also characterize the simplex structure inherent in the hyperspectral data, are used in our HU algorithm development later [2]–[8], [10], [28]:

(A1) (Non-negativity) $s_i[n] \geq 0$, $\forall\ i \in \mathcal{I}_N$ and $\forall\ n \in \mathcal{I}_L$.

(A2) (Full-additivity) $\sum_{i=1}^{N} s_i[n] = 1$, $\forall\ n \in \mathcal{I}_L$.

(A3) $\min\{L, M\} \geq N$ and $\mathbf{A} \in \mathbb{R}_+^{M \times N}$ is full column rank.

Moreover, like most benchmark HU algorithms (see, e.g., [23], [25], [28], [30]), the number of endmembers N is assumed to be known a priori, which can be determined beforehand by applying model-order selection methods, such as hyperspectral signal subspace identification by minimum error (HySiMe) [36], and Neyman-Pearson detection theory-based virtual dimensionality (VD) [37].

We aim to blindly estimate the unknown endmembers (i.e., $\mathbf{a}_1, \ldots, \mathbf{a}_N$), as well as their abundances (i.e., $\mathbf{s}[1], \ldots, \mathbf{s}[L]$), from the observed spectral mixtures (i.e., $\mathbf{x}[1], \ldots, \mathbf{x}[L]$). Due to the huge dimensionality M of hyperspectral data, directly analyzing the data may not be very computationally efficient. Instead, an efficient data preprocessing technique, called affine set fitting (ASF) procedure [38], can be applied to compactly represent each measured pixel $\mathbf{x}[n] \in \mathbb{R}^M$ in a dimension-reduced (DR) space $\mathbb{R}^{N-1}$ as follows:

$$\tilde{\mathbf{x}}[n] = \mathbf{C}^\dagger(\mathbf{x}[n] - \mathbf{d}) = \sum_{i=1}^{N} s_i[n]\boldsymbol{\alpha}_i \in \mathbb{R}^{N-1}, \tag{2}$$

---

[1]Note that there is a research line considering non-linear mixtures for modeling the effect of multiple reflections of solar radiation [34]. Moreover, the endmember spectral signatures may be spatially varying, hence leading to the full-additivity in (A2) being violated [10]. However, studying these effects is out of the scope of this paper, and the representative LMM is sufficient for our analysis and algorithm development; interested readers are referred to the magazine papers [34] and [35], respectively, for the non-linear effect and the endmember variability effect.

where

$$\mathbf{d} = \frac{1}{L} \sum_{n=1}^{L} \mathbf{x}[n] \in \mathbb{R}^M \text{ (mean of data set)} \tag{3}$$

$$\mathbf{C} = [\ \boldsymbol{q}_1(\mathbf{U}\mathbf{U}^T), \ldots, \boldsymbol{q}_{N-1}(\mathbf{U}\mathbf{U}^T)\ ] \in \mathbb{R}^{M \times (N-1)} \tag{4}$$

$$\boldsymbol{\alpha}_i = \mathbf{C}^\dagger (\mathbf{a}_i - \mathbf{d}) \in \mathbb{R}^{N-1} \text{ (DR endmembers)} \tag{5}$$

in which $\boldsymbol{q}_i(\mathbf{U}\mathbf{U}^T) \in \mathbb{R}^M$ denotes the ith principal eigenvector (with unit norm) of the square matrix $\mathbf{U}\mathbf{U}^T \in \mathbb{R}^{M \times M}$, and $\mathbf{U} = [\ \mathbf{x}[1] - \mathbf{d}, \ldots, \mathbf{x}[L] - \mathbf{d}\ ] \in \mathbb{R}^{M \times L}$ is the mean-removed data matrix. Actually, like other dimension reduction algorithms [39], ASF also performs noise suppression in the meantime. It has been shown that the above ASF best represents the measured data in an $(N-1)$-dimensional space in the sense of least-squares fitting error [38], while such fitting error vanishes in the noiseless scenario [38]. Note that the data mean in the DR space is the origin $\mathbf{0}_{N-1}$ (by (2) and (3)).

Because of $N - 1 \ll M$ in typical HU applications, the HyperCSI algorithm will be developed in the DR space $\mathbb{R}^{N-1}$ wherein the DR endmembers $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$ are estimated. Then, by (5), the endmember estimates in the original space $\mathbb{R}^M$ can be restored as

$$\hat{\mathbf{a}}_i = \mathbf{C}\ \hat{\boldsymbol{\alpha}}_i + \mathbf{d}, \quad \forall\ i \in \mathcal{I}_N, \tag{6}$$

where $\hat{\boldsymbol{\alpha}}_i$'s are the endmember estimates in the DR space.

III. HYPERPLANE-BASED CRAIG-SIMPLEX-IDENTIFICATION ALGORITHM

First of all, due to (2) and (A1)-(A2), the true endmembers' convex hull conv$\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ itself is a data-enclosing simplex, i.e.,

$$\mathcal{X} \triangleq \{\ \tilde{\mathbf{x}}[1], \ldots, \tilde{\mathbf{x}}[L]\ \} \subseteq \text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}. \tag{7}$$

According to Craig's criterion, the true endmembers' convex hull is estimated by minimizing the volume of the data-enclosing simplex [22], namely, by solving the following volume minimization problem (called the CSI problem interchangeably hereafter):

$$\min_{\boldsymbol{\beta}_i \in \mathbb{R}^{N-1}, \forall i} \quad V(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)$$

$$\text{s.t.} \quad \tilde{\mathbf{x}}[n] \in \text{conv}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}, \quad \forall n, \tag{8}$$

where $V(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)$ denotes the volume of the simplex $\mathrm{conv}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\} \subseteq \mathbb{R}^{N-1}$. Under some mild conditions on data purity level [29], the optimal solution of problem (8) can perfectly yield the true endmembers $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$. [2]

Besides in the HU context, the NP-hard CSI problem in (8) [31] has been studied in some earlier works in mathematical geology [40] and computational geometry [41]. However, their intractable computational complexity almost disable them from practical applications for larger problem size [41], mainly owing to calculation of the complicated *nonconvex* objective function [28]

$$V(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N) = \frac{1}{(N-1)!} \cdot \left| \det\left(\begin{bmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \right|$$

in (8). Instead, the HyperCSI algorithm to be presented can judiciously bypass simplex volume calculations, and meanwhile the identified simplex can be shown to be exactly the "minimum-volume" (data-enclosing) simplex in the asymptotic sense ($L \to \infty$).

First of all, let us succinctly present the actual idea on which the HyperCSI algorithm is based. As the Craig's minimum-volume simplex can be uniquely determined by N tightly enclosed $(N-2)$-dimensional hyperplanes, where each hyperplane can be reconstructed from $N-1$ affinely independent points on itself, we hence endeavor to search for $N-1$ affinely independent pixels (referred to as *active* pixels in $\mathcal{X}$) that are as close to the associated hyperplane as possible. We begin with N purest pixels that define N disjoint proper regions, each centered at a different purest pixel. Then for each hyperplane of the minimum-volume simplex, the desired $N-1$ active pixels, that are as close to the hyperplane as possible, are respectively sifted from $N-1$ subsets of $\mathcal{X}$, each enclosed in one different proper region (cf. Figure 2). Then the obtained $N-1$ pixels are used to construct one estimated hyperplane. Finally, the desired minimum-volume simplex can be determined from the obtained N hyperplane estimates.

A. Hyperplane Representation for Craig's Simplex

The idea of solving the CSI problem in (8), without involving any simplex volume computations, is based on the hyperplane representation of a simplest simplex as stated in the following proposition:

---

[2]In [29], $\widetilde{\gamma} \triangleq \max\{r' \leq 1 \mid \mathcal{T}_e \cap \mathcal{B}(r') \subseteq \mathrm{conv}\{\mathbf{s}[1], \ldots, \mathbf{s}[L]\}\}$ is used to measure the data purity level of $\mathcal{X}$, where $\mathcal{T}_e \triangleq \mathrm{conv}\{\mathbf{e}_1, \ldots, \mathbf{e}_N\} \subseteq \mathbb{R}^N$ and $\mathcal{B}(r') \triangleq \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x}\| \leq r'\}$; the geometric interpretations of $\widetilde{\gamma}$ can be found in [29]. Simply speaking, one can show that $\widetilde{\gamma} \in [\frac{1}{\sqrt{N}}, 1]$, and the most heavily mixed scenario (i.e., $\mathbf{s}[n] = \frac{1}{N}\mathbf{1}_N$, $\forall n \in \mathcal{I}_L$) will lead to the lower bound [29]. On the contrary, the pure-pixel assumption is equivalent to the condition of $\widetilde{\gamma} = 1$ (the upper bound) [29], comparing to which a mild condition of only $\widetilde{\gamma} > \frac{1}{\sqrt{N-1}}$ is sufficient to guarantee the perfect endmember identifiability of problem (8) [29].
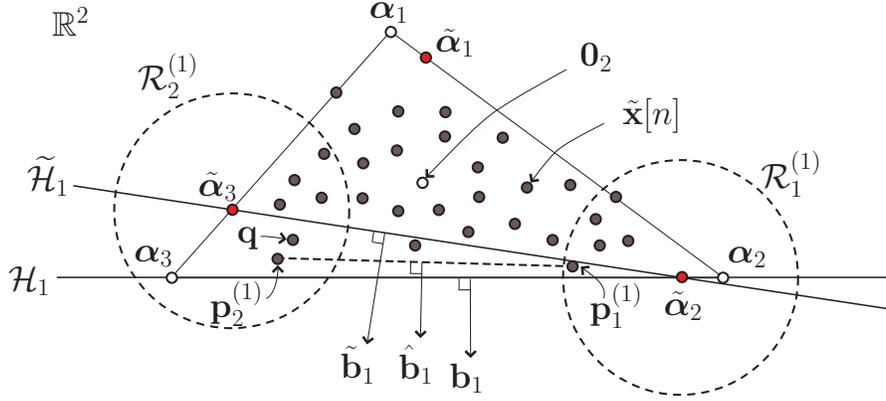
Fig. 2. An illustration of hyperplanes and DR data in $\mathbb{R}^2$ for the case of $N = 3$, where $\tilde{\boldsymbol{\alpha}}_3$ is a purest pixel in $\mathcal{X}$ (a purest pixel $\tilde{\boldsymbol{\alpha}}_i$ can be considered as the pixel closest to $\boldsymbol{\alpha}_i$) but not necessarily very close to hyperplane $\mathcal{H}_1 = \mathsf{aff}\{\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3\}$, leading to nontrivial orientation difference between $\tilde{\mathbf{b}}_1$ and $\mathbf{b}_1$. However, the active pixels $\mathbf{p}_1^{(1)}$ and $\mathbf{p}_2^{(1)}$ identified by (17) will be very close to $\mathcal{H}_1$ (especially, for large L), and hence the orientations of $\hat{\mathbf{b}}_1$ and $\mathbf{b}_1$ will be almost the same. On the other hand, one can see that the pixels identified by (21) are $\{\mathbf{p}_2^{(1)}, \mathbf{q}\}$ (that are very close to each other) whose corresponding normal vector estimate is obviously far away from the true $\mathbf{b}_1$.

Proposition 1 *If $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subseteq \mathbb{R}^{N-1}$ is affinely independent, i.e., $\mathcal{T} = \mathsf{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subseteq \mathbb{R}^{N-1}$ is a simplest simplex, then $\mathcal{T}$ can be reconstructed from the associated N hyperplanes $\{\mathcal{H}_1, \ldots, \mathcal{H}_N\}$, that tightly enclose $\mathcal{T}$, where $\mathcal{H}_i \triangleq \mathsf{aff}(\ \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \setminus \{\boldsymbol{\alpha}_i\}\ )$.*

Proof: It suffices to show that $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ can be determined by $\{\mathcal{H}_1, \ldots, \mathcal{H}_N\}$. It is known that hyperplane $\mathcal{H}_i$ can be parameterized by a normal vector $\mathbf{b}_i \in \mathbb{R}^{N-1}$ and an inner product constant $h_i \in \mathbb{R}$ as follows:

$$\mathcal{H}_i(\mathbf{b}_i, h_i) = \left\{\ \mathbf{x} \in \mathbb{R}^{N-1}\ \middle|\ \mathbf{b}_i^T \mathbf{x} = h_i\ \right\}. \tag{9}$$

As $\boldsymbol{\alpha}_i \in \mathsf{aff}(\ \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \setminus \{\boldsymbol{\alpha}_j\}\ ) = \mathcal{H}_j$ for all $j \neq i$, we have from (9) that $\mathbf{b}_j^T \boldsymbol{\alpha}_i = h_j$ for all $j \neq i$, i.e.,

$$\mathbf{B}_{-i}\boldsymbol{\alpha}_i = \mathbf{h}_{-i}, \tag{10}$$

where $\mathbf{B}_{-i} \in \mathbb{R}^{(N-1) \times (N-1)}$, $\mathbf{h}_{-i} \in \mathbb{R}^{(N-1)}$ are defined as

$$\mathbf{B}_{-i} \triangleq [\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \ldots, \mathbf{b}_N]^T, \tag{11}$$

$$\mathbf{h}_{-i} \triangleq [h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_N]^T. \tag{12}$$

As $\mathcal{T}$ is a simplest simplex in $\mathbb{R}^{N-1}$, $\mathbf{B}_{-i}$ must be of full rank and hence invertible [26]. Hence, we have from (10) that

$$\boldsymbol{\alpha}_i = \mathbf{B}_{-i}^{-1}\ \mathbf{h}_{-i},\ \forall\ i \in \mathcal{I}_N, \tag{13}$$

implying that $\boldsymbol{\alpha}_i$ can be reconstructed. The proof is therefore completed. ∎

As it can be inferred from (A3) that the set of DR endmembers $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ is affinely independent, one can apply Proposition 1 to decouple the CSI problem (8) into N subproblems of hyperplane estimation, namely, estimation of N parameter vectors $(\mathbf{b}_i, h_i)$ in (9). Then (13) can be utilized to obtain the desired endmember estimates. Next, let us present how to estimate the normal vector $\mathbf{b}_i$ and the inner product constant h_i from the data set $\mathcal{X}$, respectively.

B. Normal Vector Estimation

The normal vector $\mathbf{b}_i$ of hyperplane $\mathcal{H}_i$ can be obtained by projecting the vector $\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_i$ (for any $j \neq i$) onto the subspace that is orthogonal to the hyperplane $\mathcal{H}_i$ [42], i.e.,

$$\mathbf{b}_i \equiv \boldsymbol{v}_i(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) \tag{14}$$

$$\triangleq \left(\mathbf{I}_{N-1} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right) \cdot (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_i), \text{ for any } j \neq i,$$

where $\mathbf{P} \triangleq \mathbf{Q} - \boldsymbol{\alpha}_j \cdot \mathbf{1}_{N-2}^T \in \mathbb{R}^{(N-1)\times(N-2)}$, and $\mathbf{Q} \in \mathbb{R}^{(N-1)\times(N-2)}$ is the matrix $[\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_N] \in \mathbb{R}^{(N-1)\times N}$ with its ith and jth columns removed. Besides (14) for obtaining the normal vector $\mathbf{b}_i$ of $\mathcal{H}_i$, we also need another closed-form expression of $\mathbf{b}_i$ in terms of $N-1$ distinct points as given in the following proposition.

Proposition 2 *Given any affinely independent set* $\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\} \subseteq \mathcal{H}_i$, $\mathbf{b}_i$ *can be alternatively obtained by (except for a positive scale factor)*

$$\mathbf{b}_i = \boldsymbol{v}_i(\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{i-1}^{(i)}, \mathbf{0}_{N-1}, \mathbf{p}_i^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}), \tag{15}$$

where $\boldsymbol{v}_i(\cdot)$ *is defined in* (14).

The proof of Proposition 2 can be shown from the fact that $\mathbf{0}_{N-1}$ is the data mean in the DR space $\mathbb{R}^{N-1}$ (by (2) and (3)), and is omitted here due to space limitation.

Based on Proposition 2, we estimate the normal vector $\mathbf{b}_i$ by finding $N-1$ affinely independent data points

$$\mathcal{P}_i \triangleq \{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\} \subseteq \mathcal{X}$$

that are as close to $\mathcal{H}_i$ as possible. To this end, an observation from (7) is needed and given in the following fact:

Fact 1 *Observing that (i)* $\mathbf{b}_i^T\mathbf{p} \leq h_i$, $\forall \mathbf{p} \in \mathcal{X}$ *(i.e., all the points* $\mathbf{p} \in \mathcal{X}$ *lie on the same side of* $\mathcal{H}_i$; *cf.* (7)), *and that (ii)* $\text{dist}(\mathbf{p}, \mathcal{H}_i) = |h_i - \mathbf{b}_i^T\mathbf{p}|/\|\mathbf{b}_i\|$, *the point* $\mathbf{p} \in \mathcal{X}$ *closest to* $\mathcal{H}_i$ *is exactly the one with*

maximum of $\mathbf{b}_i^T \mathbf{p}$, provided that $\mathbf{b}_i$ points outward from the true endmembers' simplex (cf. Figure 2 and (14)).

Suppose that we are given N "purest" pixels $\tilde{\boldsymbol{\alpha}}_i \in \mathcal{X}$, which basically maximize the simplex volume inscribed in $\mathcal{X}$, and they can be obtained using the reliable and reproducible successive projection algorithm (SPA) [10], [43, Algorithm 4]. So $\tilde{\boldsymbol{\alpha}}_i$ can be viewed as the pixel in $\mathcal{X}$ "closest" to $\boldsymbol{\alpha}_i$ (cf. Figure 2). Let $\tilde{\mathbf{b}}_i$ be the outward-pointing normal vector of hyperplane $\widetilde{\mathcal{H}}_i \triangleq \text{aff}(\ \{\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_N\} \setminus \{\tilde{\boldsymbol{\alpha}}_i\}\ )$, i.e.,

$$\tilde{\mathbf{b}}_i \triangleq \boldsymbol{v}_i(\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_N). \quad \text{(cf. (14))} \tag{16}$$

Considering Fact 1 and the requirement that the set $\mathcal{P}_i$ must contain $N-1$ distinct elements (otherwise, $\mathcal{P}_i$ is not affinely independent), we identify the desired affinely independent set $\mathcal{P}_i$ by:

$$\mathbf{p}_k^{(i)} \ \in \ \arg\max \ \{\tilde{\mathbf{b}}_i^T \mathbf{p} \mid \mathbf{p} \in \mathcal{X} \cap \mathcal{R}_k^{(i)}\}, \quad \forall \ k \in \mathcal{I}_{N-1}, \tag{17}$$

where $\mathcal{R}_1^{(i)}, \dots, \mathcal{R}_{N-1}^{(i)}$ are $N-1$ disjoint sets defined as

$$\mathcal{R}_k^{(i)} \equiv \mathcal{R}_k^{(i)}(\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_N) \triangleq \begin{cases} \mathcal{B}(\tilde{\boldsymbol{\alpha}}_k, r), \ k < i, \\ \mathcal{B}(\tilde{\boldsymbol{\alpha}}_{k+1}, r), \ k \geq i, \end{cases} \tag{18}$$

in which $\mathcal{B}(\tilde{\boldsymbol{\alpha}}_k, r) \triangleq \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x} - \tilde{\boldsymbol{\alpha}}_k\| < r\}$ is the open Euclidean norm ball with center $\tilde{\boldsymbol{\alpha}}_k \in \mathbb{R}^N$ and radius $r \triangleq (1/2) \cdot \min\{\|\tilde{\boldsymbol{\alpha}}_i - \tilde{\boldsymbol{\alpha}}_j\| \mid 1 \leq i < j \leq N\} > 0$. Note that the choice of the radius r is to guarantee that $\mathcal{R}_1^{(i)}, \dots, \mathcal{R}_{N-1}^{(i)}$ are $N-1$ non-overlapping regions, thereby guaranteeing that $\mathcal{P}_i$ contains $N-1$ distinct points. Moreover, each hyperball $\mathcal{R}_k^{(i)}$ must contain at least one pixel (as it contains either $\tilde{\boldsymbol{\alpha}}_k$ or $\tilde{\boldsymbol{\alpha}}_{k+1}$; cf. (18)), i.e., $\mathcal{X} \cap \mathcal{R}_k^{(i)} \neq \emptyset$, and hence problem (17) must be a feasible problem (i.e., a problem with non-empty feasible set [26]).

If the $N-1$ points extracted by (17) are affinely independent, then the estimated normal vector associated with $\mathcal{H}_i$ can be determined as (cf. Proposition 2)

$$\hat{\mathbf{b}}_i = \boldsymbol{v}_i(\mathbf{p}_1^{(i)}, \dots, \mathbf{p}_{i-1}^{(i)}, \mathbf{0}_{N-1}, \mathbf{p}_i^{(i)}, \dots, \mathbf{p}_{N-1}^{(i)}). \tag{19}$$

Fortunately, the obtained $\mathcal{P}_i$ by (17) can be proved (in Theorem 1 below) to be always affinely independent with one more assumption: [3]

(A4) The abundance vectors $\{\mathbf{s}[n]\} \subseteq \mathbb{R}^N$ (defined below (1)) are independent and identically distributed (i.i.d.) following Dirichlet distribution with parameter $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_N] \in \mathbb{R}^N_{++}$ whose probability density function (p.d.f.) is given by [44]:

$$f(\mathbf{s}) = \begin{cases} \frac{\Gamma(\gamma_0)}{\prod_{i=1}^N \Gamma(\gamma_i)} \cdot \prod_{i=1}^N s_i^{\gamma_i - 1}, & \mathbf{s} \in \mathsf{dom}\ f, \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

where $\mathbf{s} = [s_1, \ldots, s_N] \in \mathbb{R}^N$, $\gamma_0 = \sum_{i=1}^N \gamma_i$, $\mathsf{dom}\ f = \{\mathbf{s} \in \mathbb{R}^N_{++} \mid \mathbf{1}_N^T \mathbf{s} = 1\}$, and $\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x}\ dx$ denotes the gamma function.

Theorem 1 *Assume* (A1)-(A4) *hold true. Let* $\mathbf{p}_k^{(i)} \in \mathcal{P}_i$ *be a solution to* (17) *with* $\mathcal{R}_k^{(i)}$ *defined in* (18), *for all* $i \in \mathcal{I}_N$ *and* $k \in \mathcal{I}_{N-1}$. *Then, the set* $\mathcal{P}_i$ *is affinely independent with probability* 1 *(w.p.1).*

The proof of Theorem 1 is given in Appendix A.

Note that the orientation difference between $\tilde{\mathbf{b}}_i$ and the true $\mathbf{b}_i$ may not be small (cf. Figure 2). Hence, $\tilde{\mathbf{b}}_i$ itself may not be a good estimate for $\mathbf{b}_i$ either. On the contrary, it can be shown that the orientation difference between $\hat{\mathbf{b}}_i$ and $\mathbf{b}_i$ tends to be small for large L, and actually such difference vanishes as L goes to infinity (cf. Theorem 2 as well as Remark 1 in Subsection III-E). On the other hand, if the pixels with maximum inner products in $\mathcal{P}_i$ are jointly sifted from the whole data cloud $\mathcal{X}$, i.e.,

$$\mathcal{P}_i \in \arg\max\ \{\tilde{\mathbf{b}}_i^T(\mathbf{p}_1 + \cdots + \mathbf{p}_{N-1}) \mid \mathcal{P} \subseteq \mathcal{X}\}, \tag{21}$$

where $\mathcal{P} \triangleq \{\mathbf{p}_1, \ldots, \mathbf{p}_{N-1}\}$, rather than respectively from different regions $\mathcal{X} \cap \mathcal{R}_k^{(i)}$, $\forall k \in \mathcal{I}_{N-1}$, as given in (17), the identified pixels in $\mathcal{P}_i$ may stay quite close, easily leading to large deviation in normal vector estimation as illustrated in Figure 2 where $\mathcal{P}_i = \{\mathbf{p}_2^{(1)}, \mathbf{q}\}$ are the identified pixels using (21). This is also a rationale of finding $\mathcal{P}_i$ using (17) for better normal vector estimation.

---

[3]The rationale of adopting Dirichlet distribution in (A4) is not only that it is a well known distribution that captures both the non-negativity and full-additivity of $\mathbf{s}[n]$ [44], but because it has been used to characterize the distribution of $\mathbf{s}[n]$ in the HU context [45], [46]. However, the statistical assumption (A4) is only for analysis purpose without being involved in our geometry-oriented algorithm development. So even if abundance vectors are neither i.i.d. nor Dirichlet distributed, the HyperCSI algorithm can still work well; cf. Subsection IV-D. Furthermore, we would like to emphasize that, in our analysis (Theorems 1 and 2), we actually only use the following two properties of Dirichlet distribution: (i) its domain is $\mathsf{dom}\ f = \{\mathbf{s} \in \mathbb{R}^N_{++} \mid \mathbf{1}_N^T \mathbf{s} = 1\}$, and (ii) it is a continuous multivariate distribution with strictly positive density on its domain [47]; cf. Appendixes A and B. Hence, any distribution with these two properties can be used as an alternative in (A4).

C. Inner Product Constant Estimation

For Craig's simplex (the minimum-volume data-enclosing simplex), all the data in $\mathcal{X}$ should lie on the same side of $\mathcal{H}_i$ (otherwise, it is not data-enclosing), and $\mathcal{H}_i$ should be as tightly close to the data cloud $\mathcal{X}$ as possible (otherwise, it is not minimum-volume); the only possibility is when the hyperplane $\mathcal{H}_i$ must be externally tangent to the data cloud. In other words, $\mathcal{H}_i$ will incorporate the pixel that has maximum inner product with $\hat{\mathbf{b}}_i$, and hence it can be determined as $\mathcal{H}_i(\hat{\mathbf{b}}_i, \hat{h}_i)$, where $\hat{h}_i$ is obtained by solving

$$\hat{h}_i = \max\ \{\ \hat{\mathbf{b}}_i^T \mathbf{p} \mid \mathbf{p} \in \mathcal{X}\ \}. \tag{22}$$

However, it has been reported that when the observed data pixels are noise-corrupted, the random noise may expand the data cloud, thereby inflating the volume of the Craig's data-enclosing simplex [21], [33]. As a result, the estimated hyperplanes are pushed away from the origin (i.e., the data mean in the DR space) due to noise effect, and hence the estimated inner product constant in (22) would be larger than that of the ground truth. To mitigate this effect, the estimated hyperplanes need to be properly shifted closer to the origin, so instead, $\mathcal{H}_i(\hat{\mathbf{b}}_i, \hat{h}_i/c)$, $\forall\ i \in \mathcal{I}_N$, are the desired hyperplane estimates for some $c \geq 1$. Therefore, the corresponding DR endmember estimates are obtained by (cf. (13))

$$\hat{\boldsymbol{\alpha}}_i = \widehat{\mathbf{B}}_{-i}^{-1} \cdot \frac{\hat{\mathbf{h}}_{-i}}{c}, \quad \forall\ i \in \mathcal{I}_N, \tag{23}$$

where $\widehat{\mathbf{B}}_{-i}$ and $\hat{\mathbf{h}}_{-i}$ are given by (11) and (12) with $\mathbf{b}_j$ and h_j replaced by $\hat{\mathbf{b}}_j$ and $\hat{h}_j$, $\forall\ j \neq i$, respectively. Moreover, it is necessary to choose c such that the associated endmember estimates in the original space are non-negative (cf. (A3)), i.e.,

$$\hat{\mathbf{a}}_i = \mathbf{C}\ \hat{\boldsymbol{\alpha}}_i + \mathbf{d} \succeq \mathbf{0}_M,\ \forall\ i \in \mathcal{I}_N.\ \ (\text{cf. (6)}) \tag{24}$$

By (23) and (24), the hyperplanes should be shifted closer to the origin with $c = c'$ at least, where

$$c' \triangleq \min_{c'' \geq 1} \{c'' \mid \mathbf{C}\ (\widehat{\mathbf{B}}_{-i}^{-1} \cdot \hat{\mathbf{h}}_{-i}) + c'' \cdot \mathbf{d} \succeq \mathbf{0}_M,\ \forall\ i\} \tag{25}$$

which can be further shown to have a closed-form solution:

$$c' = \max\ \{1, \max\{-v_{ij}/d_j \mid i \in \mathcal{I}_N,\ j \in \mathcal{I}_M\}\}, \tag{26}$$

where v_{ij} is the jth component of $\mathbf{C}\ (\widehat{\mathbf{B}}_{-i}^{-1} \cdot \hat{\mathbf{h}}_{-i}) \in \mathbb{R}^M$ and d_j is the jth component of $\mathbf{d}$.

Note that c' is just the minimum value for c to yield non-negative endmember estimates. Thus, we can generally set $c = c'/\eta \geq c'$ for some $\eta \in (0, 1]$. Moreover, the value of $\eta = 0.9$ is empirically found to be a good choice for the scenarios where signal-to-noise ratio (SNR) is greater than 20 dB; typically, the

value of SNR in hyperspectral data is much higher than 20 dB [32]. Let us emphasize that the larger the value of η (or the smaller the value of c), the farther the estimated hyperplanes from the origin $\mathbf{0}_{N-1}$, or the closer the estimated endmembers' simplex $\text{conv}\{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_N\}$ to the boundary of the nonnegative orthant $\mathbb{R}_+^M$. On the other hand, we empirically observed that typical endmembers in the U.S. geological survey (USGS) library [48] are close to the boundary of $\mathbb{R}_+^M$. That is to say, a reasonable choice of $\eta \in (0, 1]$ should be relatively large (i.e., relatively close to 1), accounting for the reason why the preset value of $\eta = 0.9$ can always yield good performance. The resulting endmember estimation processing of the HyperCSI algorithm is summarized in Steps 1 to 6 in Table I.

TABLE I

PSEUDO-CODE FOR HYPERCSI ALGORITHM

Given	Hyperspectral data $\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}$, number of endmembers N, and $\eta = 0.9$.
Step 1.	Calculate $(\mathbf{C}, \mathbf{d})$ using (3)-(4), and obtain the DR data $\mathcal{X} = \{\tilde{\mathbf{x}}[1], \ldots, \tilde{\mathbf{x}}[L]\}$ using (2).
Step 2.	Obtain $\{\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N\}$ using SPA [43].
Step 3.	Obtain $\tilde{\mathbf{b}}_i$ using (16), $\forall\, i$, and $\mathcal{R}_k^{(i)}$ using (18), $\forall\, i, k$.
Step 4.	Obtain $(\mathcal{P}_i, \hat{\mathbf{b}}_i, \hat{h}_i)$ by (17), (19), and (22), $\forall\, i$.
Step 5.	Obtain c' by (26), and set $c = c'/\eta$.
Step 6.	Calculate $\hat{\boldsymbol{\alpha}}_i$ by (23) and $\hat{\mathbf{a}}_i = \mathbf{C}\, \hat{\boldsymbol{\alpha}}_i + \mathbf{d}$ by (24), $\forall\, i$.
Step 7.	Calculate $\hat{\mathbf{s}}[n] = [\hat{s}_1[n] \cdots \hat{s}_N[n]]^T$ by (30), $\forall\, n$.
Output	The endmember estimates $\{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_N\}$ and abundance estimates $\{\hat{\mathbf{s}}[1], \ldots, \hat{\mathbf{s}}[L]\}$.

D. Abundance Estimation

Though the abundance estimation is often done by solving FCLS problems [15], which can be equivalently formulated in the DR space as (cf. [49, Lemma 3.1])

$$
\min_{\mathbf{s}' \in \mathbb{R}^N} \quad \|\tilde{\mathbf{x}}[n] - [\hat{\boldsymbol{\alpha}}_1 \cdots \hat{\boldsymbol{\alpha}}_N]\mathbf{s}'\|
$$
$$
\text{s.t.} \quad \mathbf{s}' \succeq \mathbf{0}_N, \ \mathbf{1}_N^T \mathbf{s}' = 1, \tag{27}
$$

it has been reported that some geometric quantities, acquired during the endmember extraction stage, can be used to significantly accelerate the abundance estimation procedure [50]. With similar computational efficiency improvements taken into account, we aim at expressing the abundance $s_i[n]$ in terms of readily available quantities (e.g., normal vectors and inner product constants) obtained when estimating the endmembers, in this subsection. The results are summarized in the following proposition:

Proposition 3 *Assume (A1)-(A3) hold true. Then* $\mathbf{s}[n] = [s_1[n] \cdots s_N[n]]^T$ *has the following closed-form expression:*

$$s_i[n] = \frac{h_i - \mathbf{b}_i^T \tilde{\mathbf{x}}[n]}{h_i - \mathbf{b}_i^T \boldsymbol{\alpha}_i}, \quad \forall \, i \in \mathcal{I}_N, \ \forall \, n \in \mathcal{I}_L. \tag{28}$$

Proposition 3 can be derived from some simple geometrical observations (cf. items (i) and (ii) in Fact 1) and the following well known formula in the *Algebraic Topology* context

$$s_i[n] = \frac{\text{dist}(\tilde{\mathbf{x}}[n], \mathcal{H}_i)}{\text{dist}(\boldsymbol{\alpha}_i, \mathcal{H}_i)}, \tag{29}$$

and its proof is omitted here due to space limitation; note that the formula (29) has been recently derived again using different approach in the HU context [50, Equation (12)].

Based on (28), the abundance vector $\mathbf{s}[n]$ can be estimated as

$$\hat{s}_i[n] = \left(\frac{\hat{h}_i - \hat{\mathbf{b}}_i^T \tilde{\mathbf{x}}[n]}{\hat{h}_i - \hat{\mathbf{b}}_i^T \hat{\boldsymbol{\alpha}}_i} \right)^+, \quad \forall \, i \in \mathcal{I}_N, \ \forall \, n \in \mathcal{I}_L, \tag{30}$$

where $(y)^+ \triangleq \max\{y, 0\}$ is to enforce the non-negativity of abundance fractions $\mathbf{s}[n]$ (cf. Step 7 in Table I). One can show that when $\tilde{\mathbf{x}}[n] \in \text{conv}\{\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_N\}$, the abundance estimates obtained using (30) is exactly the solution to the FCLS problem in (27), while using (30) has much lower computational cost than solving FCLS problems. Nevertheless, one should be aware of a potential limitation of using (30). Specifically, if $\tilde{\mathbf{x}}[n]$ is too far away from the endmembers' simplex $\text{conv}\{\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_N\}$ (i.e., $\hat{\mathbf{b}}_i^T \tilde{\mathbf{x}}[n]$ is much larger than $\hat{h}_i$ for some i), the zeroing operation in (30) could cause nontrivial deviation in abundance estimation. This can happen if $\tilde{\mathbf{x}}[n]$ is an outlier or the SNR is very low. However, as the SNR is reasonably high (like in AVIRIS data [21], [32]), most pixels in the hyperspectral data are expected to lie inside or very close to the endmembers' simplex $\text{conv}\{\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_N\}$ (cf. (A1)-(A2))—especially when the endmembers are extracted based on Craig's criterion. Hence, with the endmembers estimated by the Craig-criterion-based HyperCSI algorithm, simply using (30) to enforce the abundance non-negativity is not only computationally efficient, but also still capable of yielding good abundance estimation as will be demonstrated in the simulation results (Table III in Subsection IV-C and Table IV in Subsection IV-D) later.

Unlike most of the existing abundance estimation algorithms, where all the N abundance maps must be jointly estimated (e.g., FCLS [15]), the proposed HyperCSI algorithm offers an option of solely obtaining the abundance map of a specific material of interest (say the ith material)

$$\boldsymbol{s}_i \triangleq [s_i[1] \cdots s_i[L]]^T \in \mathbb{R}_+^L, \tag{31}$$

to save computational cost, or obtaining all the abundance maps $s_1, \ldots, s_N$ by parallel processing (cf. (30)). Moreover, when calculating s_i using (30), the denominator $\hat{h}_i - \hat{\mathbf{b}}_i^T \hat{\boldsymbol{\alpha}}_i$ is a constant for all pixel indices $n \in \mathcal{I}_L$ and hence only needs to be calculated once regardless of L (which is usually large).

E. Identifiability and Complexity of HyperCSI

In this subsection, let us present the identifiability and complexity analyses of the proposed HyperCSI algorithm. Particularly, the asymptotic identifiability of the HyperCSI algorithm can be guaranteed as stated in the following theorem with the proof given in Appendix B:

Theorem 2 *Under (A1)-(A4), the noiseless assumption and $L \to \infty$, the simplex identified by HyperCSI algorithm with $c = 1$ (in Step 5 in Table I) is exactly the Craig's minimum-volume simplex (i.e., solution of* (8)*) and the true endmembers' simplex* $\mathsf{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ *w.p.1.*

Two noteworthy remarks about the philosophies and intuitions behind the proof of this theorem are given as follows:

Remark 1 With the abundance distribution stated in (A4), the $N - 1$ pixels in $\mathcal{P}_i$ can be shown to be arbitrarily close to $\mathcal{H}_i$ as the pixel number $L \to \infty$, and they are affinely independent w.p.1 (cf. Theorem 1). Therefore, $\hat{\mathbf{b}}_i$ can be uniquely obtained by (19), and its orientation approaches to that of $\mathbf{b}_i$ w.p.1.

Remark 2 Remark 1 together with (7) implies that $\hat{h}_i$ is upper bounded by h_i w.p.1 (assuming without loss of generality that $\|\hat{\mathbf{b}}_i\| = \|\mathbf{b}_i\|$), and this upper bound can be shown to be achievable w.p.1 as $L \to \infty$. Thus, as $c = 1$, we have that $\hat{h}_i/c = h_i$ w.p.1.

It can be further inferred, from the above two remarks, that $\hat{\boldsymbol{\alpha}}_i$ is exactly the true $\boldsymbol{\alpha}_i$ w.p.1 (cf. (23)) as $L \to \infty$ in the absence of noise. Although the identifiability analysis in Theorem 2 is conducted for the noiseless case and $L \to \infty$, we empirically found that the HyperCSI algorithm can yield good endmember estimates for a moderate L and finite SNR, to be demonstrated by simulation results and real data experiments later.

Next, we analyze the computational complexity of the HyperCSI algorithm. The computation time of HyperCSI is primarily dominated by the computations of the feasible sets $\mathcal{X} \cap \mathcal{R}_k^{(i)}$ (in Step 3), the active pixels in $\mathcal{P}_i$ (in Step 4), and the abundances $\hat{\mathbf{s}}[n]$ (in Step 7), and they are respectively analyzed in the following:

Step 3: Computing the $N(N-1)$ feasible sets $\mathcal{X} \cap \mathcal{R}_k^{(i)}$, $\forall i \in \mathcal{I}_N$, $\forall k \in \mathcal{I}_{N-1}$, is equivalent to computing the N sets $\mathcal{X} \cap \mathcal{B}(\tilde{\boldsymbol{\alpha}}_i, r)$, $\forall i \in \mathcal{I}_N$; cf. (18). Since $\mathcal{B}(\tilde{\boldsymbol{\alpha}}_i, r)$ is an open Euclidean norm ball, the

computation of each set $\mathcal{X} \cap \mathcal{B}(\tilde{\boldsymbol{\alpha}}_i, r)$ can be done by examining $|\mathcal{X}| = L$ inequalities $\|\tilde{\mathbf{x}}[n] - \tilde{\boldsymbol{\alpha}}_i\| < r$, $\forall n \in \mathcal{I}_L$. However, examining each inequality requires (i) calculating one Euclidean 2-norm (in $\mathbb{R}^{N-1}$), which costs $\mathcal{O}(N)$, and (ii) checking whether this 2-norm is smaller than r, which costs $\mathcal{O}(1)$. Hence, Step 3 costs $\mathcal{O}(N(N+1)L)$.

Step 4: To determine $\mathcal{P}_i$, we have to identify the pixel $\mathbf{p}_k^{(i)}$ from the set $\mathcal{X} \cap \mathcal{R}_k^{(i)}$ ($\forall k \in \mathcal{I}_{N-1}$), whose complexity amounts to computing $|\mathcal{X} \cap \mathcal{R}_k^{(i)}|$ inner products in $\mathbb{R}^{N-1}$ (each costs $\mathcal{O}(N)$), and performing the point-wise maximum operation among the values of these inner products (cf. (17)), and hence the complexity of identifying $\mathbf{p}_k^{(i)}$ is easily verified as $\mathcal{O}(N \cdot |\mathcal{X} \cap \mathcal{R}_k^{(i)}| + |\mathcal{X} \cap \mathcal{R}_k^{(i)}| - 1)$. Moreover, gathering $\mathcal{P}_i = \{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\}$ requires the complexity $\sum_{k=1}^{N-1} \mathcal{O}(N \cdot |\mathcal{X} \cap \mathcal{R}_k^{(i)}| + |\mathcal{X} \cap \mathcal{R}_k^{(i)}| - 1) = \mathcal{O}((N+1) \cdot \sum_{k=1}^{N-1} |\mathcal{X} \cap \mathcal{R}_k^{(i)}|) \leq \mathcal{O}((N+1) \cdot |\mathcal{X}|) = \mathcal{O}((N+1)L)$; the inequality is due to that $\mathcal{R}_k^{(i)}$'s are disjoint. Repeating the above for $\mathcal{P}_i$, $\forall i \in \mathcal{I}_N$, Step 4 costs $\mathcal{O}(N(N+1)L)$.

Step 7: Estimation of the abundances requires to compute the fraction in (30) NL times. Each fraction involves 2 inner products (in $\mathbb{R}^{N-1}$), 2 scalar subtractions, and 1 scalar division, and thus costs $\mathcal{O}(N)$. So, this step costs $\mathcal{O}(N^2 L)$.

Therefore, the overall computational complexity of HyperCSI is $\mathcal{O}(2N(N+1)L + N^2 L) = \mathcal{O}(N^2 L)$.

Surprisingly, the complexity order $\mathcal{O}(N^2 L)$ of the proposed HyperCSI algorithm is the same as (rather than much higher than) that of some pure-pixel-based EEAs; see, e.g., [17], [21], [43], [51]. Moreover, to the best of our knowledge, the MVES algorithm [28] that approximates the CSI problem in (8) as alternating linear programming (LP) problems, and solves the LPs using primal-dual interior-point method [26], is the existing Craig-criterion-based algorithm with lowest complexity order $\mathcal{O}(\tau N^2 L^{1.5})$, where τ is the number of iterations [28]. Hence, the introduced hyperplane identification approach (without simplex volume computations) indeed yields a smaller complexity than most existing Craig-criterion-based algorithms.

Let us conclude this section with a summary of some remarkable features of the proposed HyperCSI algorithm (given in Table I) as follows:

(a) Without involving any simplex volume computations, the Craig's minimum-volume simplex is reconstructed from N hyperplane estimates, i.e., the N estimates $(\hat{\mathbf{b}}_i, \hat{h}_i)$, which can be obtained in parallel (cf. Step 4 in Table I) by searching $N(N-1)$ most informative pixels from $\mathcal{X}$. The reconstructed simplex in the DR space $\mathbb{R}^{N-1}$ is actually the intersection of N halfspaces $\{\mathbf{x} \in \mathbb{R}^{N-1} \mid \hat{\mathbf{b}}_i^T \mathbf{x} \leq \hat{h}_i\}$, $i = 1, \ldots, N$.

(b) By noting that $s_i[n] = 0$ if, and only if, $\mathbf{x}[n] \in \mathcal{H}_i$, the potential requirement of $N(N-1)$ pixels lying on, or close to, the associated hyperplanes is considered not difficult to be met in practice because

hyperspectral images are often with highly sparse abundances. This will be discussed in more detail in experiments with AVIRIS data in Section V.

(c) All the processing steps (including SPA in Step 2 of Table I; cf. Algorithm 4 in [43]) can be carried out either by simple linear algebraic formulations or by closed-form expressions, and so its high computational efficiency can be anticipated.

IV. COMPUTER SIMULATIONS

This section demonstrates the efficacy of the proposed HyperCSI algorithm by Monte Carlo simulations. In the simulation, endmember signatures with $M = 224$ spectral bands randomly selected from the USGS library [48] are used to generate L noise-free synthetic hyperspectral data $\mathbf{x}[n]$ according to linear mixing model in (1), where the abundance vectors are i.i.d. and generated following the Dirichlet distribution with $\boldsymbol{\gamma} = \mathbf{1}_N/N$ (cf. (20)) as it can automatically enforce (A1) and (A2) [28], [33]. Then we add i.i.d. zero-mean Gaussian noise with variance σ^2 to the noise-free synthetic data $\mathbf{x}[n]$ for different values of SNR defined as $\text{SNR} \triangleq (\sum_{n=1}^{L} \|\mathbf{x}[n]\|^2)/(\sigma^2 ML)$, and those negative entries in the generated noisy data vectors are artificially set to zero, so as to maintain the non-negativity nature of hyperspectral imaging data.

The root-mean-square (rms) spectral angle error between the true endmembers $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ and their estimates $\{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_N\}$ defined as [17], [28]

$$\phi_{en} = \min_{\boldsymbol{\pi} \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\arccos \left(\frac{\mathbf{a}_i^T \hat{\mathbf{a}}_{\pi_i}}{\|\mathbf{a}_i\| \cdot \|\hat{\mathbf{a}}_{\pi_i}\|} \right) \right]^2}, \tag{32}$$

is used as the performance measure of endmember estimation, where $\Pi_N = \{ \boldsymbol{\pi} = (\pi_1, \ldots, \pi_N) \in \mathbb{R}^N \mid \pi_i \in \{1, \ldots, N\}, \ \pi_i \neq \pi_j \text{ for } i \neq j \}$ is the set of all permutations of $\{1, \ldots, N\}$. Similarly, the performance measure of abundance estimation is the rms angle error defined as [28]

$$\phi_{ab} = \min_{\boldsymbol{\pi} \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\arccos \left(\frac{\boldsymbol{s}_i^T \hat{\boldsymbol{s}}_{\pi_i}}{\|\boldsymbol{s}_i\| \cdot \|\hat{\boldsymbol{s}}_{\pi_i}\|} \right) \right]^2}, \tag{33}$$

where $\boldsymbol{s}_i$ and $\hat{\boldsymbol{s}}_i$ are the true abundance map of ith endmember (cf. (31)) and its estimate, respectively. All the HU algorithms under test are implemented in Mathworks Matlab R2013a running on a desktop computer equipped with Core-i7-4790K CPU with 4.00 GHz speed and 16 GB random access memory, and all the performance results in terms of ϕ_{en}, ϕ_{ab}, and computational time T are averaged over 100 independent realizations.

Next, we show some simulation results for the endmember identifiability for moderately finite data length (cf. Theorem 2), the choice of the parameter η, and the performance evaluation of the proposed HyperCSI algorithm, in the following subsections, respectively.
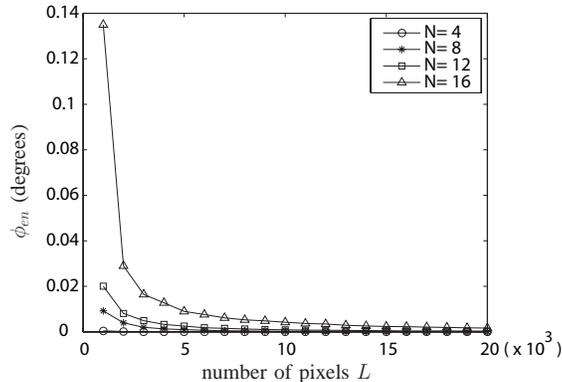
Fig. 3. The endmember identifiability of the HyperCSI algorithm with finite data length L.

A. Endmember Identifiability of HyperCSI for Finite Data

In Theorem 2, the perfect endmember identifiability of the proposed HyperCSI algorithm (with $c = 1$ in Step 5 in Table I) under the noise-free scenario is proved in the asymptotic sense (i.e., the data length $L \to \infty$). In this subsection, we would like to show some simulation results to illustrate the asymptotic identifiability of the HyperCSI algorithm and its good endmember estimation accuracy even with a moderately finite number of pixels L.

Figure 3 shows some simulation results of ϕ_{en} versus L for $N \in \{4, 8, 12, 16\}$. From this figure, one can observe that for a given N, ϕ_{en} decreases as L increases, and the HyperCSI algorithm indeed achieves perfect identifiability (i.e., $\phi_{en} = 0$, cf. (32)) as $L \to \infty$. On the other hand, the HyperCSI algorithm needs to identify $N(N-1)$ essential pixels $\mathbf{p}_k^{(i)}$ for the construction of the Craig's simplex, which indicates that the HyperCSI algorithm would need more pixels to achieve good performance for larger N. Intriguingly, the results shown in Figure 3 are consistent with the above inferences, where a larger N corresponds to a slightly slower convergence rate of ϕ_{en}. However, these results also allude to a high possibility that the HyperCSI algorithm can yield accurate endmember estimates with a typical data length L (i.e., several ten thousands) for high SNR in HRS applications.

B. Choice of the Parameter η

The simulation results for ϕ_{en} versus η obtained by the proposed HyperCSI algorithm, for $L = 10000$, $\text{SNR} \in \{20, 30, 40\}$ (dB), and $N \in \{3, 4, 5, 6\}$ are shown in Figure 4. From this figure, one can observe that for a fixed N, the best choice of η (i.e., the one that yields the smallest ϕ_{en}) decreases as SNR decreases. The reason for this is that the larger the noise power, the more the data cloud is expanded, and hence the more the desired hyperplanes should be shifted towards the data center (implying a larger
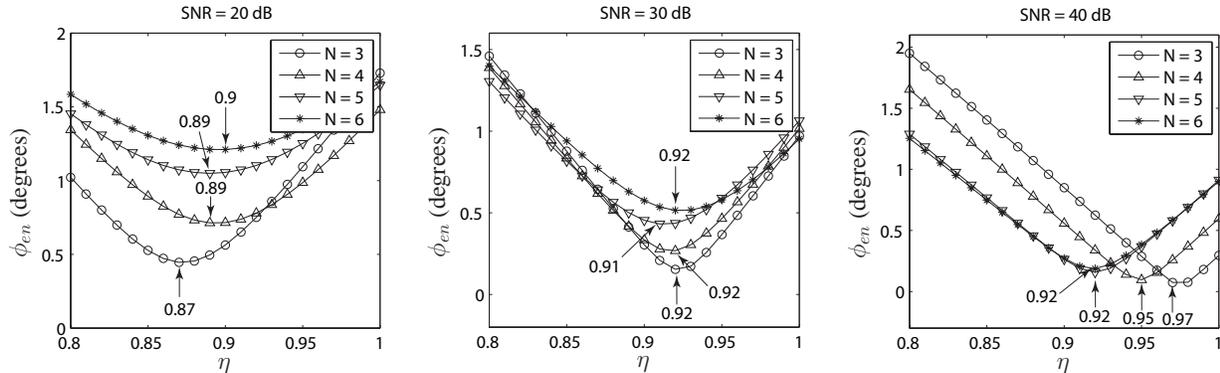
Fig. 4. The average r.m.s. spectral angle error ϕ_{en} versus different values of η.

c or a smaller η). Moreover, one can also observe from Figure 4 that for each scenario of (N, SNR), the best choice of η basically belongs to the interval $[0.8, 1]$, a relatively large value in the interval $(0, 1]$, as discussed in Subsection III-C. It is also interesting to note that for a given SNR, the best choice of η tends to approach the value of 0.9 as N increases. For instance, for $\text{SNR} = 20$ dB, the best choices of η for $N \in \{3, 4, 5, 6\}$ are $\{0.87, 0.89, 0.89, 0.9\}$, respectively.

The above observations also suggest that $\eta = 0.9$, the only parameter in the proposed HyperCSI algorithm, is a good choice. Next, we will evaluate the performance of the proposed HyperCSI algorithm with the parameter η preset to 0.9 for all the simulated scenarios and real data tests, though it may not be the best choice for some scenarios.

C. Performance Evaluation of HyperCSI Algorithm

We evaluate the performance of the proposed HyperCSI algorithm, along with a performance comparison with five state-of-the-art Craig-criterion-based HU algorithms, including MVC-NMF [23], MVSA [25], MVES [28], SISAL [30], and ipMVSA [27]. As the operations of MVC-NMF, MVSA, SISAL, and ipMVSA are data-dependent, their respective regularization parameters have been well selected in the simulation, so as to yield their best performances. In particular, the regularization parameter involved in SISAL is the regression weight for robustness against noise, and hence has also been tuned w.r.t. different SNRs. The implementation details and parameter settings for all the algorithms under test are listed in Table II.

The purity index ρ_n for each synthetic pixel $\mathbf{x}[n]$ [28], [29], [33] has been defined as $\rho_n \triangleq \|\mathbf{s}[n]\| \in [1/\sqrt{N}, 1]$ (due to (A1) and (A2)); a larger index ρ_n means higher pixel purity of $\mathbf{x}[n] = \sum_{i=1}^{N} s_i[n] \mathbf{a}_i$. Each synthetic data set in the simulation is generated with a given purity level denoted as ρ, following

TABLE II

SIMULATION SETTINGS FOR THE ALGORITHMS UNDER TEST.

Algorithms	Implementation details and parameter settings
MVC-NMF	Dimension reduction: Singular value decomposition; Regularization parameter: 10^{-3}; Max iteration: 500; Initialization: VCA-FCLS; Convergence tolerance: 10^{-6}.
MVSA	Dimension reduction: Principal component analysis; Regularization parameter: 10^{-6}; Initialization: VCA.
MVES	Dimension reduction: ASF; Convergence tolerance: 10^{-8}; Initialization: Solving feasibility problem.
SISAL	Dimension reduction: Principal component analysis; Regularization parameter: $0.015, 0.02, 0.025, 0.03, 0.035$ w.r.t SNR= $20, 25, 30, 35, 40$ (dB); Initialization: VCA.
ipMVSA	Dimension reduction: Principal component analysis; Regularization parameter: 10^{-6}; Initialization: VCA.
HyperCSI	Dimension reduction: ASF; $\eta = 0.9$.

the same data generation procedure as in [28], [29], [33], where ρ is a measure of mixing level of a data set. Specifically, a pool of sufficiently large number of synthetic data is first generated, and then from the pool, L pixels with the purity index ρ_n not greater than ρ are randomly picked to form the desired data set with a purity level of ρ.

In the above data generation, six endmembers (i.e., Jarsoite, Pyrope, Dumortierite, Buddingtonite, Muscovite, and Goethite) with 224 spectral bands randomly selected from the USGS library [48] are used to generate 10000 synthetic hyperspectral data $\mathbf{x}[n]$ (i.e., $N = 6$, $M = 224$, $L = 10000$) with $\rho \in \{0.8, 0.9, 1\}$ and SNR $\in \{20, 25, 30, 35, 40\}$ (dB). The simulation results for ϕ_{en}, ϕ_{ab}, and computational time T are displayed in Table III, where bold-face numbers correspond to the best performance (i.e., the smallest ϕ_{en}, ϕ_{ab}, and T) of all the HU algorithms under test for a specific (ρ, SNR).

Some general observations from Table III are as follows. For fixed purity level ρ, all the algorithms under test perform better for larger SNR. As expected, the proposed HyperCSI algorithm rightly performs better for higher data purity level ρ, but this performance behavior does not apply to the other five algorithms, perhaps because the non-convexity of the complicated simplex volume makes their performance behaviors more intractable w.r.t. different data purities.

Among the five existing benchmark Craig-criterion-based HU algorithms, MVC-NMF yields more

TABLE III

PERFORMANCE COMPARISON, IN TERMS OF ϕ_{en} (DEGREES), ϕ_{ab} (DEGREES), AND AVERAGE RUNNING TIME T (SECONDS), OF VARIOUS HU ALGORITHMS FOR DIFFERENT DATA PURITY LEVELS ρ AND SNRS, WHERE ABUNDANCES ARE I.I.D. AND DIRICHLET DISTRIBUTED.

Methods	ρ	ϕ_{en} (degrees) SNR (dB)					ϕ_{ab} (degrees) SNR (dB)					T (seconds)
		20	25	30	35	40	20	25	30	35	40	
MVC-NMF	0.8	2.87	2.31	1.63	1.23	1.14	13.18	9.83	7.14	5.58	5.04	1.68E+2
	0.9	2.98	1.78	0.98	0.57	0.40	12.67	8.00	4.64	2.85	2.16	
	1	3.25	1.91	1.00	0.52	**0.21**	12.30	7.45	4.14	2.26	**1.11**	
MVSA	0.8	11.08	6.23	3.41	1.87	1.03	21.78	14.49	8.71	5.00	2.85	3.54E+0
	0.9	11.55	6.46	3.48	1.90	1.05	21.89	14.51	8.63	4.91	2.82	
	1	11.64	6.51	3.54	1.93	1.06	21.67	14.21	8.49	4.81	2.72	
MVES	0.8	10.66	6.06	3.39	1.91	1.16	21.04	14.21	9.04	5.51	3.33	2.80E+1
	0.9	10.17	6.06	3.48	1.97	1.12	21.51	14.48	9.28	5.69	3.45	
	1	9.95	5.96	3.55	2.19	1.30	22.50	15.34	10.32	7.11	4.49	
SISAL	0.8	3.97	2.59	1.59	0.94	0.53	13.70	8.68	5.22	3.09	1.80	2.59E+0
	0.9	4.18	2.70	1.64	0.95	0.54	13.55	8.54	5.11	3.00	1.75	
	1	4.49	2.87	1.73	0.99	0.54	13.40	8.43	5.03	2.93	1.66	
ipMVSA	0.8	12.03	7.05	4.04	2.02	1.16	21.81	14.89	9.58	5.32	2.23	9.86E-1
	0.9	12.63	7.55	4.04	2.05	1.25	22.33	15.36	9.37	5.21	3.31	
	1	12.89	7.80	4.00	2.13	1.28	22.16	15.20	9.06	5.25	3.28	
HyperCSI	0.8	**1.65**	**1.20**	**0.79**	**0.54**	**0.37**	**11.17**	**7.35**	**4.32**	**2.65**	**1.64**	**5.39E-2**
	0.9	**1.37**	**1.03**	**0.64**	**0.45**	**0.32**	**10.08**	**6.40**	**3.62**	**2.25**	**1.38**	
	1	**1.21**	**0.83**	**0.57**	**0.39**	0.27	**9.28**	**5.46**	**3.23**	**1.92**	1.15	

accurate endmember estimates than the other algorithms at the highest computational cost, while ipMVSA is the most computationally efficient one with lower performance as a trade-off. Nevertheless, the proposed HyperCSI algorithm outperforms all the other five algorithms when the data are heavily mixed (i.e., $\rho = 0.8$) or moderately mixed (i.e., $\rho = 0.9$). As for high data purity $\rho = 1$, the HyperCSI algorithm also performs best except for the case of $(\rho, \mathrm{SNR}) = (1, 40 \text{ dB})$. On the other hand, the computational efficiency of the proposed HyperCSI algorithm is about 1 to 4 orders of magnitude faster than the other five HU algorithms under test. Note that the computational efficiency of the HyperCSI algorithm can be further improved by an order of $\mathcal{O}(N)$ if parallel processing can be implemented in Step 4 (hyperplane estimation) and Step 7 (abundance estimation) in Table I. Moreover, ipMVSA is around 4 times faster than MVSA, but performs slightly worse than MVSA, perhaps because ipMVSA [27] does not adopt the hinge-type soft constraint (for noise resistance) as used in MVSA [25].

D. Performance Evaluation of HyperCSI Algorithm with Non-i.i.d., Non-Dirichlet and Sparse Abundances

In practice, the abundance vectors $\mathbf{s}[n]$ may not be i.i.d. and seldom follow the Dirichlet distribution, and, moreover, the abundance maps often show large sparseness [52]. In view of this, as considered in [52], [53], two sets of sparse and spatially correlated abundance maps displayed in Figure 5 were used to generate two synthetic hyperspectral images, denoted as SYN1 ($L = 100 \times 100$) and SYN2 ($L = 130 \times 130$). Then all the algorithms listed in Table II are tested again with these two synthetic data sets for which the abundance vectors are obviously neither i.i.d. nor Dirichlet distributed.

(a) Ground truth abundance maps of SYN1

(b) Ground truth abundance maps of SYN2
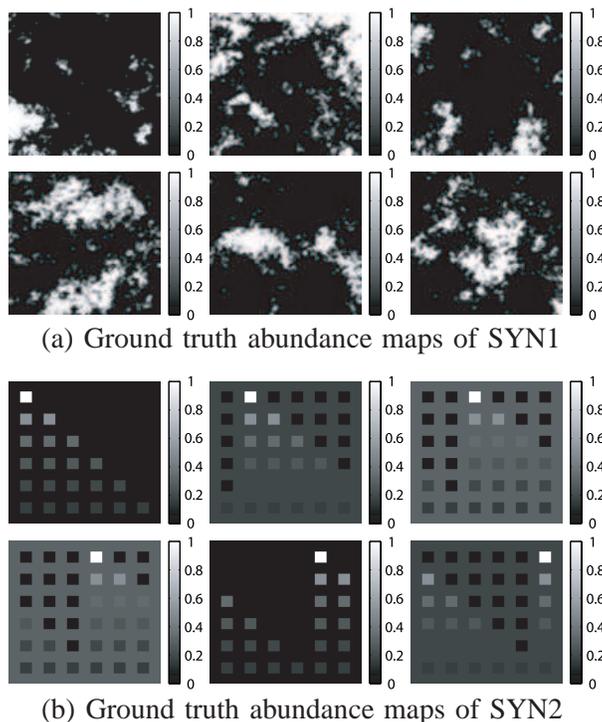
Fig. 5. Two sets of sparse and spatially correlated abundance maps, where each subblock in subfigure (b) contains 10×10 pixels.

The simulation results, in terms of ϕ_{en}, ϕ_{ab}, and computational time T, are shown in Table IV, where bold-face numbers correspond to the best performance among the algorithms under test for a particular data set and a specific SNR $\in \{20, 25, 30, 35, 40\}$ (dB). As expected, for both data sets, all the algorithms perform better for larger SNR.

One can see from Table IV that for both data sets, HyperCSI yields more accurate endmember estimates than the other algorithms, except for the case of SNR= 40 (dB). As for abundance estimation, HyperCSI performs best for SYN1, while MVC-NMF performs best for SYN2. Moreover, among the five existing

TABLE IV

PERFORMANCE COMPARISON, IN TERMS OF ϕ_{en} (DEGREES), ϕ_{ab} (DEGREES) AND RUNNING TIME T (SECONDS), OF VARIOUS HU ALGORITHMS USING SYNTHETIC DATA SYN1 AND SYN2 FOR DIFFERENT SNRS, WHERE ABUNDANCES ARE NON-I.I.D., NON-DIRICHLET AND SPARSE (SEE FIGURE 5).

	Methods	ϕ_{en} (degrees)					ϕ_{ab} (degrees)					T (seconds)
		SNR (dB)					SNR (dB)					
		20	25	30	35	40	20	25	30	35	40	
SYN1	MVC-NMF	3.23	1.97	1.05	0.55	**0.25**	13.87	8.51	4.79	2.65	**1.34**	1.74E+2
	MVSA	10.65	6.12	3.38	1.88	1.05	22.93	15.13	9.34	5.52	3.19	3.53E+0
	MVES	9.55	5.49	3.60	1.96	1.22	23.89	17.35	14.49	7.78	5.66	3.42E+1
	SISAL	4.43	2.89	1.81	1.18	0.86	15.85	10.39	6.89	5.29	4.65	2.66E+0
	ipMVSA	11.62	6.82	3.38	2.01	1.05	24.05	16.28	9.34	5.98	3.19	1.65E+0
	HyperCSI	**1.55**	**1.22**	**0.79**	**0.52**	0.35	**12.03**	**6.92**	**4.16**	**2.49**	1.46	**5.56E-2**
SYN2	MVC-NMF	2.86	1.71	0.97	0.54	**0.23**	22.86	**15.52**	**9.39**	**5.27**	**2.67**	2.48E+2
	MVSA	10.21	5.55	3.08	1.71	0.95	29.86	22.72	15.57	9.78	5.83	5.65E+0
	MVES	10.12	5.19	3.15	2.04	3.77	29.43	22.13	15.66	10.42	13.17	2.22E+1
	SISAL	3.25	2.18	1.48	0.96	0.63	24.79	17.49	11.51	7.00	4.21	4.45E+0
	ipMVSA	11.34	8.26	3.34	1.94	1.01	30.23	30.38	16.29	10.30	6.39	8.14E-1
	HyperCSI	**1.48**	**1.08**	**0.71**	**0.44**	0.31	**22.64**	15.98	11.10	7.25	4.40	**7.48E-2**

benchmark Craig-criterion-based HU algorithms, ipMVSA and SISAL are the most computationally efficient ones. However, in both data sets, the computational efficiency of the proposed HyperCSI algorithm is at least more than one order of magnitude faster than the other five algorithms. These simulation results have demonstrated the superior efficacy of the proposed HyperCSI algorithm over the other algorithms under test in both estimation accuracy and computational efficiency.

V. EXPERIMENTS WITH AVIRIS DATA

In this section, the proposed HyperCSI algorithm along with two benchmark HU algorithms, i.e., the MVC-NMF algorithm [23] developed based on Craig's criterion, and the VCA algorithm [17] (in conjunction with the FCLS algorithm [15] for the abundance estimation) developed based on the pure-pixel assumption, are used to process the hyperspectral imaging data collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [32] taken over the Cuprite mining site, Nevada, in 1997. We consider this mining site, not only because it has been extensively used for remote sensing experiments [54], but also because the available classification ground truth in [55], [56] (though which may have coregistration

issue as it was obtained earlier than 1997, this ground truth has been widely accepted in the HU context) allows us to easily verify the experimental results. The AVIRIS sensor is an imaging spectrometer with 224 channels (or spectral bands) that cover wavelengths ranging from 0.4 to 2.5 μm with an approximately 10-nm spectral resolution. The bands with low SNR as well as those corrupted by water-vapor absorption (including bands 1-4, 107-114, 152-170, and 215-224) are removed from the original 224-band imaging data cube, and hence a total of $M = 183$ bands is considered in our experiments. Furthermore, the selected subscene of interest includes 150 vertical lines with 150 pixels per line, and its 50th band is shown in Figure 6(a), where the 10 pixels marked with yellow color are removed from the data set as they are outlier pixels identified by the robust affine set fitting (RASF) algorithm [57].
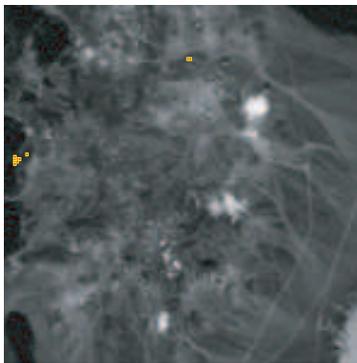
Fig. 6.   The subimage of the AVIRIS hyperspectral imaging data cube for the 50th band, where the locations of the ten outliers identified by the RASF algorithm are marked with yellow color.

The number N of the minerals (i.e., endmembers) present in the selected subscene is estimated using a virtual dimensionality (VD) approach [37], i.e., the noise-whitened Harsanyi-Farrand-Chang (NWHFC) eigenvalue-thresholding-based algorithm with false-alarm probability $P_{FA} = 10^{-3}$. The obtained estimate is $\hat{N} = 9$ and used in the ensuing experiments for all the three HU algorithms under test.

The estimated abundance maps are visually compared with those reported in [17], [23], [28] as well as the ground truth reported in [55], [56], so as to determine what minerals they are associated with. The nine abundance maps obtained by the proposed HyperCSI algorithm are shown in Figure 7, and they are identified as mineral maps of Muscovite, Alunite, Desert Varnish, Hematite, Montmorillonite, Kaolinite #1, Kaolinite #2, Buddingtonite, Chalcedony, respectively, as listed in Table V. The minerals identified by MVC-NMF and VCA are also listed in Table V, where MVC-NMF also identifies nine distinct minerals, while only eight distinct minerals are retrieved by VCA, perhaps due to lack of pure pixels in the selected subscene or randomness involved in VCA. Owing to space limitation, their mineral

maps are not shown here.
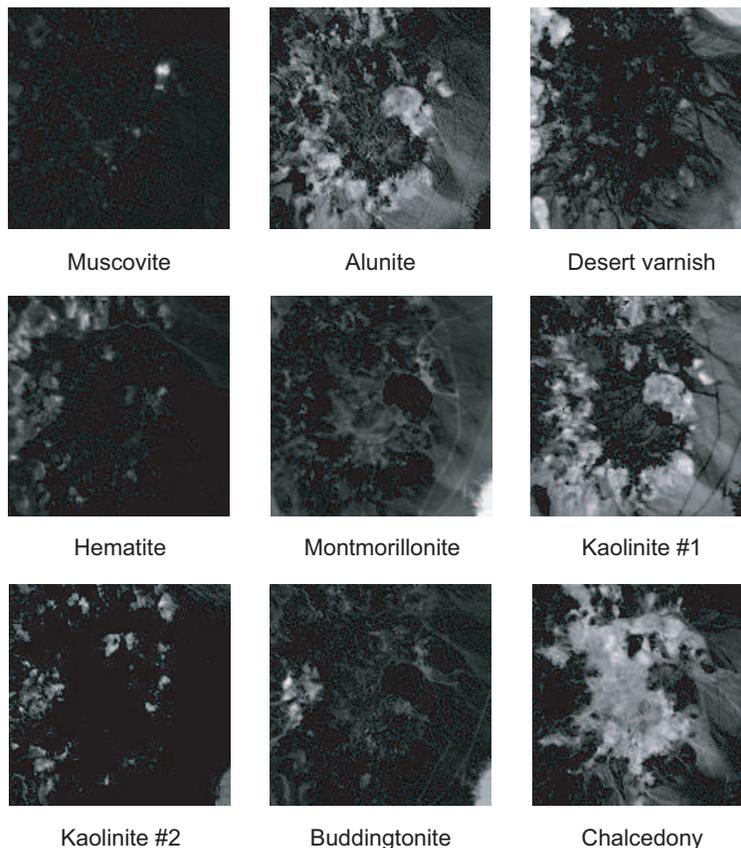
Fig. 7. The abundance maps of minerals estimated by HyperCSI algorithm.

The mineral spectra extracted by the three algorithms under test, along with their counterparts in the USGS library [48], are shown in Figure 8, where one can observe that the spectra extracted by the proposed HyperCSI algorithm hold a better resemblance to the library spectra. For instance, the spectrum of Alunite extracted by HyperCSI shows much clearer absorption feature than MVC-NMF and VCA, in the bands approximately from 2.3 to 2.5 μm. To quantitatively compare the endmember estimation accuracy among the three algorithms under test, the spectral angle distance between each endmember estimate $\hat{\mathbf{a}}$ and its corresponding library spectrum $\mathbf{a}$ serves as the performance measure and is defined as

$$\phi = \arccos\left(\frac{\mathbf{a}^T\hat{\mathbf{a}}}{\|\mathbf{a}\| \cdot \|\hat{\mathbf{a}}\|}\right).$$

The values of ϕ associated with the endmember estimates for all the three algorithms under test are also shown in Table V, where the number in the parentheses is the value of ϕ associated with Kaolinite #1

TABLE V

THE COMPUTATIONAL TIMES T (SECONDS) AND SPECTRAL ANGLE DISTANCE ϕ (DEGREES) BETWEEN LIBRARY SPECTRA AND ENDMEMBERS ESTIMATED BY HYPERCSI, MVC-NMF, AND VCA. THE BOLD FACE NUMBERS CORRESPOND TO THE SMALLEST VALUES OF ϕ OR T AMONG THE THREE ALGORITHMS UNDER TEST.

	HyperCSI	MVC-NMF	VCA
Muscovite	**3.03**	3.96	4.54
Alunite	7.48	**6.23**	6.57
Desert Varnish	9.49	**4.91**	7.92
Hematite	7.83	12.94	**7.24**
Montmorillonite	**4.84**	7.44	6.59
Kaolinite #1	8.63	**7.56**	13.80 (11.71)
Kaolinite #2	**7.39**	-	-
Buddingtonite	6.55	8.16	**6.46**
Chalcedony	**5.92**	7.97	8.25
Andradite	-	**7.43**	-
Average ϕ (degrees)	**6.80**	7.40	8.12
T (seconds)	**0.12**	988.67	5.40

repeatedly classified by VCA. One can see from Table V that the average of ϕ of the proposed HyperCSI algorithm is the smallest. The good performance of HyperCSI in endmember estimation intimates to that the potential requirement of sufficient number (i.e., $N(N-1) = 72$, in this experiment) of pixels lying close to the hyperplanes associated with the actual endmembers' simplex, has been met. However, we are not too surprised with this observation, since the number of minerals present in one pixel $\mathbf{x}[n]$ is often small (typically, within five [10]), i.e., the abundance vector $\mathbf{s}[n]$ often shows sparseness [52] (cf. Figure 7), indicating that a non-trivial portion of pixels are more likely to lie close to the boundary of the endmembers' simplex (note that $s_i[n] = 0$ if, and only if, $\mathbf{x}[n] \in \mathcal{H}_i$). Moreover, as the pure pixels may not be present in the selected subscene, as expected the two Craig-criterion-based HU algorithms (i.e., HyperCSI and MVC-NMF) outperform VCA in terms of endmember estimation accuracy. On the other hand, in terms of the computation time T as given in Table V, in spite of parallel processing not applied, the HyperCSI algorithm is around 2.5 times faster than VCA (note that VCA itself only costs 0.31 seconds (out of the 5.40 seconds), and the remaining computation time is the cost of the FCLS) and almost four orders of magnitude faster than MVC-NMF.
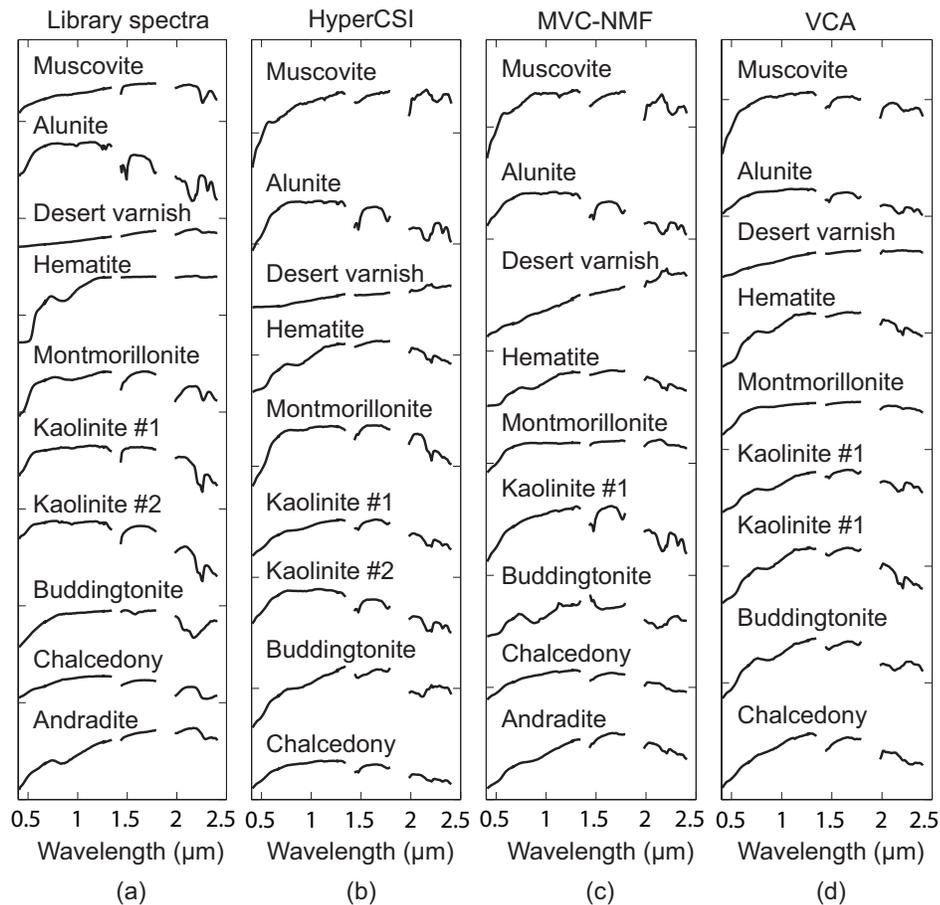
Fig. 8. (a) The endmember signatures taken from the USGS library, and signatures of the endmember estimates obtained by (b) HyperCSI, (c) MVC-NMF and (d) VCA.

VI. CONCLUSIONS

Based on the hyperplane representation for a simplest simplex, we have presented an effective and computationally efficient Craig-criterion-based HU algorithm, called HyperCSI algorithm, given in Table I. The proposed HyperCSI algorithm has the following remarkable characteristics:

- It never requires the presence of pure pixels in the data.
- It is reproducible without involving random initialization.
- It only involves simple linear algebraic computations, and suitable for parallel implementation. Its computational complexity (without using parallel implementation) is $\mathcal{O}(N^2 L)$, which is also the complexity of some state-of-the-art pure-pixel-based HU algorithms.

- It estimates Craig's minimum-volume simplex by finding only $N(N-1)$ pixels (regardless of the data length L) from the data set for the construction of the associated hyperplanes, without involving any simplex volume computations, thereby accounting for its high computational efficiency in endmember estimation.

- The estimated endmembers are guaranteed non-negative, and the identified simplex was proven to be both Craig's simplex and true endmembers' simplex w.p.1. as $L \to \infty$ for the noiseless case.

- The abundance estimation is readily fulfilled by a closed-form expression, and thus is computationally efficient.

Some simulation results were presented to demonstrate the analytic results on the asymptotic endmember identifiability of the proposed HyperCSI algorithm, and its superior efficacy over some state-of-the-art Craig-criterion-based HU algorithms in both solution accuracy and computational efficiency. Finally, the proposed HyperCSI algorithm was tested using AVIRIS hyperspectral data to show its applicability.

APPENDIX

A. Proof of Theorem 1

For a fixed $i \in \mathcal{I}_N$, one can see from (18) that $\mathcal{R}_k^{(i)} \cap \mathcal{R}_\ell^{(i)} = \emptyset$, $\forall\, k \neq \ell$, implying that the $N-1$ pixels $\mathbf{p}_k^{(i)}$, $\forall\, k \in \mathcal{I}_{N-1}$, identified by solving (17) must be distinct. Hence, it suffices to show that $\mathcal{P}$ is affinely independent w.p.1 for any $\mathcal{P} \triangleq \{\mathbf{p}_1, \ldots, \mathbf{p}_{N-1}\} \subseteq \mathcal{X}$ that satisfies

$$\mathbf{p}_k \neq \mathbf{p}_\ell, \text{ for all } 1 \leq k < \ell \leq N - 1. \tag{34}$$

Then, as $\mathbf{p}_k \in \mathcal{X}$, $\forall\, k \in \mathcal{I}_{N-1}$, we have from (A4) and (34) that there exist i.i.d Dirichlet distributed random vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\} \subseteq \text{dom } f$ such that (cf. (2))

$$\mathbf{p}_k = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_N]\, \mathbf{s}_k, \quad \text{for all } k \in \mathcal{I}_{N-1}. \tag{35}$$

For ease of the ensuing presentation, let $\Pr\{\cdot\}$ denote the probability function and define the following events:

      E1   The set $\mathcal{P}$ is affinely dependent.

      E2   The set $\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\}$ is affinely dependent.

      E3$^{(k)}$ $\mathbf{s}_k \in \text{aff}\{\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\} \setminus \{\mathbf{s}_k\}\}$, $\forall\, k \in \mathcal{I}_{N-1}$.

Then, to prove that $\mathcal{P}_i$ is affinely independent w.p.1, it suffices to prove $\Pr\{\mathsf{E1}\} = 0$.

Next, let us show that E1 implies E2. Assume E1 is true. Then $\mathbf{p}_k \in \text{aff}\{\mathcal{P} \setminus \{\mathbf{p}_k\}\}$ for some $k \in \mathcal{I}_{N-1}$. Without loss of generality, let us assume $k = 1$. Then,

$$\mathbf{p}_1 = \theta_2 \cdot \mathbf{p}_2 + \cdots + \theta_{N-1} \cdot \mathbf{p}_{N-1}, \tag{36}$$

for some $\theta_i, i = 2, \ldots, N-1$, satisfying

$$\theta_2 + \cdots + \theta_{N-1} = 1. \tag{37}$$

By substituting (35) into (36), we have

$$[\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N] \, \mathbf{s}_1 = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N] \, \mathbf{t}, \tag{38}$$

where $\mathbf{t} \triangleq \sum_{m=2}^{N-1} \theta_m \cdot \mathbf{s}_m$. For notational simplicity, let $[\mathbf{u}]_{1:N-1} \triangleq [u_1, \ldots, u_{N-1}]^T$ for any given vector $\mathbf{u} = [u_1, \ldots, u_N]^T$. Then, from the facts of $\mathbf{1}_N^T \mathbf{t} = 1$ (by (37)) and $\mathbf{1}_N^T \mathbf{s}_1 = 1$, (38) can be rewritten as

$$\boldsymbol{\Theta} \, [\mathbf{s}_1]_{1:N-1} = \boldsymbol{\Theta} \, [\mathbf{t}]_{1:N-1}, \tag{39}$$

where $\boldsymbol{\Theta} \triangleq [\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_N, \ldots, \boldsymbol{\alpha}_{N-1} - \boldsymbol{\alpha}_N]$. As $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ is affinely independent (by (A3)), the matrix $\boldsymbol{\Theta}$ is of full column rank [26], implying that $[\mathbf{s}_1]_{1:N-1} = [\mathbf{t}]_{1:N-1}$ (by (39)). Then, by the facts of $\mathbf{1}_N^T \mathbf{t} = 1$ and $\mathbf{1}_N^T \mathbf{s}_1 = 1$, one can readily come up with $\mathbf{s}_1 = \mathbf{t} = \sum_{m=2}^{N-1} \theta_m \cdot \mathbf{s}_m$, or, equivalently, $\mathbf{s}_1 \in \text{aff}\{\mathbf{s}_2, \ldots, \mathbf{s}_{N-1}\}$ (by (37)), implying that E2 is true [26]. Thus we have proved that E1 implies E2, and hence

$$\Pr\{\mathsf{E1}\} \le \Pr\{\mathsf{E2}\}. \tag{40}$$

As Dirichlet distribution is a continuous multivariate distribution [47] for a random vector $\mathbf{s} \in \mathbb{R}^N$ to satisfy (A1)-(A2) with an $(N-1)$-dimensional domain, any given affine hull $\mathcal{A} \subseteq \mathbb{R}^N$ with affine dimension P must satisfy [44]

$$\Pr\{ \mathbf{s} \in \mathcal{A} \} = 0, \text{ if } P < N - 1. \tag{41}$$

Moreover, as $\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\}$ are i.i.d. random vectors and the affine hull $\text{aff}\{\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\} \setminus \{\mathbf{s}_k\}\}$ must have affine dimension $P < N - 1$, we have from (41) that

$$\Pr\{\mathsf{E3}^{(k)}\} = 0, \text{ for all } k \in \mathcal{I}_{N-1}. \tag{42}$$

Then we have the following inferences:

$$0 \le \Pr\{\mathsf{E1}\} \le \Pr\{\mathsf{E2}\} \qquad \text{(by (40))}$$

$$= \Pr\{\cup_{k=1}^{N-1} \mathsf{E3}^{(k)}\} \qquad \text{(by the definitions of E2 and } \mathsf{E3}^{(k)})$$

$$\le \sum_{k=1}^{N-1} \Pr\{\mathsf{E3}^{(k)}\} = 0, \qquad \text{(by the union bound and (42))}$$

i.e., $\Pr\{\mathsf{E1}\} = 0$. Therefore, the proof is completed. ∎

B. Proof of Theorem 2

It can be seen from (20) that the p.d.f. of Dirichlet distribution satisfies

$$f(\mathbf{s}) = \frac{\Gamma(\gamma_0)}{\prod_{i=1}^{N} \Gamma(\gamma_i)} \cdot \prod_{i=1}^{N} s_i^{\gamma_i - 1} > 0, \ \forall \ \mathbf{s} \in \text{dom } f. \tag{43}$$

Moreover, by the facts of $\mathbf{A}\mathbf{e}_i = \mathbf{a}_i$ and

$$\text{dom } f = \{\mathbf{s} \in \mathbb{R}_{++}^N \mid \mathbf{1}_N^T \mathbf{s} = 1\} = \text{int conv}\{\mathbf{e}_1, \ldots, \mathbf{e}_N\},$$

where int $\mathcal{U}$ denotes the interior of a set $\mathcal{U}$, the linear mapping (i.e., $\mathbf{x} = \mathbf{A}\mathbf{s}$) of the abundance domain dom f full fills the interior of the true endmembers' simplex conv$\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$, namely

$$\{\mathbf{x} = \mathbf{A}\mathbf{s} \mid \mathbf{s} \in \text{dom } f\} = \text{int conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}. \tag{44}$$

Then, from (43)-(44) and (A4), it can be inferred that

$$\text{Pr}\{\text{conv}\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}|_{L \to \infty} = \text{int conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}\} = 1,$$

which, together with the fact that the affine mapping (cf. (2)) preserves the geometric structure of $\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}$ [38] (note that $\mathbf{C}^T \mathbf{C} = \mathbf{I}_{N-1}$), further implies

$$\text{Pr}\{ \text{conv}\mathcal{X} = \text{int conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \} = 1, \tag{45}$$

where $\mathcal{X} = \{\tilde{\mathbf{x}}[1], \ldots, \tilde{\mathbf{x}}[L]\}|_{L \to \infty}$ throughout the ensuing proof. It can be inferred from (45) that there is always a pixel $\tilde{\mathbf{x}}[n] \in \mathcal{X}$ that can be arbitrarily close to the extreme point $\boldsymbol{\alpha}_i$ of the simplex conv$\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$, i.e., for all $i \in \mathcal{I}_N$,

$$\text{Pr}\{ \mathcal{B}(\boldsymbol{\alpha}_i, \epsilon) \cap \mathcal{X} \neq \emptyset \} = 1, \ \text{for any } \epsilon > 0. \tag{46}$$

Let MVES($\mathcal{U}$) denote the set of all minimum-volume enclosing simplexes of $\mathcal{U}$ (i.e., Craig's simplex containing the set $\mathcal{U}$). Then, one can infer from the convexity of a simplex that (cf. [29, Equation (32)])

$$\text{MVES}(\mathcal{X}) = \text{MVES}(\text{conv}\mathcal{X}). \tag{47}$$

Moreover, by the fact that any simplex $\mathcal{T}$ must also be a closed set and the fact that the closure of int(conv$\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$) is exactly conv$\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$, it can be seen that conv$\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subseteq \mathcal{T}$ if and only if int(conv$\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$) $\subseteq \mathcal{T}$ [58], and hence

$$\text{MVES}(\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}) = \text{MVES}(\text{int conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}). \tag{48}$$

Thus, it can be inferred from (45), (47) and (48) that

$$\text{Pr}\{ \text{MVES}(\mathcal{X}) = \text{MVES}(\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}) \} = 1. \tag{49}$$

As $\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ itself is a simplex, $\text{MVES}(\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}) = \{\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}\}$, which together with (49) yields

$$\text{Pr}\{\ \text{MVES}(\mathcal{X}) = \{\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}\}\ \} = 1. \tag{50}$$

In other words, we have proved that the Craig's minimum-volume simplex is always the true endmembers' simplex $\text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$. To complete the proof of Theorem 2, it suffices to show that the true endmembers' simplex is always identical to the simplex identified by the HyperCSI algorithm, i.e., for all $i \in \mathcal{I}_N$,

$$\text{Pr}\{\ \hat{\boldsymbol{\alpha}}_i \in \mathcal{B}(\boldsymbol{\alpha}_i, \epsilon)\ \} = 1, \ \text{for any } \epsilon > 0, \tag{51}$$

where $\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_N$ are the estimated DR endmembers using HyperCSI algorithm.

To this end, let us first show that, for all $i \in \mathcal{I}_N$,

$$\text{Pr}\{\ \tilde{\boldsymbol{\alpha}}_i \in \mathcal{B}(\boldsymbol{\alpha}_i, \epsilon)\ \} = 1, \ \text{for any } \epsilon > 0, \tag{52}$$

where $\{\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N\}$ are the purest pixels identified by SPA (cf. Step 2 in Table I). However, directly proving (52) is difficult due to the post-processing involved in SPA (cf. Algorithm 4 in [43]). In view of this, let $\tilde{\mathbf{x}}[\ell_1], \ldots, \tilde{\mathbf{x}}[\ell_N]$ be those pixels identified by SPA *before* post-processing. Because the post-processing is nothing but to obtain the purest pixel $\tilde{\boldsymbol{\alpha}}_i$ by iteratively pushing each $\tilde{\mathbf{x}}[\ell_i]$ away from the hyperplane $\text{aff}\{\tilde{\mathbf{x}}[\ell_j] \mid j \neq i\}$ [43], we have the following simplex volume inequalities

$$V(\tilde{\mathbf{x}}[\ell_1], \ldots, \tilde{\mathbf{x}}[\ell_N]) \leq V(\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N) \leq V(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N), \tag{53}$$

where the last inequality is due to $\tilde{\boldsymbol{\alpha}}_i \in \mathcal{X} \subseteq \text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$. Hence, by (53), to prove (52), it suffices to show that, for all $i \in \mathcal{I}_N$,

$$\text{Pr}\{\ \tilde{\mathbf{x}}[\ell_i] \in \mathcal{B}(\boldsymbol{\alpha}_i, \epsilon)\ \} = 1, \ \text{for any } \epsilon > 0. \tag{54}$$

However, the SPA before post-processing (cf. Algorithm 4 in [43]) is exactly the same as the TRIP algorithm (cf. Algorithm 2 in [51]), and it has been proven in [51, Lemma 3] that (46) straightforwardly yields (54) for $\epsilon = 0$; note that the condition "(46) with $\epsilon = 0$" is equivalent to the pure-pixel assumption required in [51, Lemma 3]. One can also show that (46) yields (54) for any $\epsilon > 0$, and the proof basically follows the same induction procedure as in the proof of [51, Lemma 3] and is omitted here for conciseness. Then, recalling that (54) is a sufficient condition for (52) to hold, we have proven (52).

By the fact that $\boldsymbol{v}_i$ is a continuous function (cf. (14)) and by (16) and (18), we see that

$$\begin{aligned}
\lim_{\tilde{\boldsymbol{\alpha}}_i \to \boldsymbol{\alpha}_i, \ \forall i} \tilde{\mathbf{b}}_i &= \boldsymbol{v}_i(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) = \mathbf{b}_i, \\
\lim_{\tilde{\boldsymbol{\alpha}}_i \to \boldsymbol{\alpha}_i, \ \forall i} \mathcal{R}_k^{(i)} &= \mathcal{R}_k^{(i)}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N).
\end{aligned} \tag{55}$$

Moreover, we have from (43), (52), (55) and $L \to \infty$ that the pixel $\mathbf{p}_k^{(i)}$ identified by (17) can be arbitrarily close to $\mathcal{H}_i$. Furthermore, by Theorem 1, we have that the vectors $\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\}$ are not only arbitrarily close to $\mathcal{H}_i$, but also affinely independent w.p.1, which together with Proposition 2 implies that the estimated $\hat{\mathbf{b}}_i$ (cf. (19)) is arbitrarily close to the true $\mathbf{b}_i$ w.p.1, provided that the outward-pointing normal vectors $\hat{\mathbf{b}}_i$ and $\mathbf{b}_i$ have the same norm without loss of generality. Then, from (43), (22), and the premises of $L \to \infty$ and $c = 1$, it can be inferred that the estimated hyperplane $\hat{\mathcal{H}}_i \equiv \mathcal{H}_i(\hat{\mathbf{b}}_i, \hat{h}_i/c) = \mathcal{H}_i(\hat{\mathbf{b}}_i, \hat{h}_i)$ is arbitrarily close to the true $\mathcal{H}_i \equiv \mathcal{H}_i(\mathbf{b}_i, h_i)$ (cf. (9)); precisely, we have

$$\Pr\{ \ [\hat{\mathbf{b}}_i^T, \hat{h}_i]^T \in \mathcal{B}([\mathbf{b}_i^T, h_i]^T, \epsilon) \ \} = 1, \ \text{for any } \epsilon > 0. \tag{56}$$

Consequently, by comparing the formulas of $\boldsymbol{\alpha}_i$ (cf. (13)) and $\hat{\boldsymbol{\alpha}}_i$ (cf. (23)), we have, from $c = 1$ and (56), that $\hat{\boldsymbol{\alpha}}_i$ is always arbitrarily close to $\boldsymbol{\alpha}_i$, i.e., (51) is true for all $i \in \mathcal{I}_N$, and hence the proof of Theorem 2 is completed. ∎

REFERENCES

[1] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based MVES algorithm for hyperspectral unmixing," in *Proc. IEEE ICASSP*, Brisbane, Australia, Apr. 19-24, 2015, pp. 1384–1388.

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

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

[4] W.-K. Ma, J. M. Bioucas-Dias, J. Chanussot, and P. Gader, Eds., *Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing, IEEE Signal Process. Mag.*, vol. 31, no. 1, Jan. 2014.

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

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

[7] 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.

[8] J. 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.*, vol. 5, no. 2, pp. 354–379, 2012.

[9] W. R. Johnson, M. Humayun, G. Bearman, D. W. Wilson, and W. Fink, "Snapshot hyperspectral imaging in ophthalmology," *Journal of Biomedical Optics*, vol. 12, no. 1, pp. 0 140 361–0 140 367, Feb. 2007.

[10] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, 2014.

[11] A. Zymnis, S.-J. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Proc. 41st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 4-7, 2007.

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

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

[14] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 446–450, 2007.

[15] 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.

[16] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summ. JPL Airborne Earth Sci. Workshop*, vol. 1, Pasadena, CA, Dec. 9-14, 1995, pp. 23–26.

[17] 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.

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

[19] 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, 2006.

[20] C.-I. Chang, C.-C. Wu, C. S. Lo, and M.-L. Chang, "Real-time simplex growing algorithms for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1834–1850, 2010.

[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, 2011.

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

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

[24] E. M. Hendrix, I. García, J. Plaza, G. Martin, and A. Plaza, "A new minimum-volume enclosing algorithm for endmember identification and abundance estimation in hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2744–2757, 2012.

[25] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, vol. 4, Boston, MA, Aug. 8-12, 2008, pp. 2369–2371.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[27] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Apr. 2015.

[28] 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, 2009.

[29] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no pure-pixel case," to appear in *IEEE Trans. Geosci. Remote Sens.*, 2015, available on http://arxiv.org/abs/1406.5273.

[30] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE WHISPERS*, Grenoble, France, Aug. 26-28, 2009, pp. 1–4.

[31] A. Packer, "NP-hardness of largest contained and smallest containing simplices for V- and H-polytopes," *Discrete and Computational Geometry*, vol. 28, no. 3, pp. 349–377, 2002.

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

[33] 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, 2011.

[34] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. Bermudez, S. Mclaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, 2014.

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

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

[37] 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.

[38] 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.

[39] I. K. Fodor, "A survey of dimension reduction techniques," *Tech. Rep. UCRL-ID-148494, Lawrence Livermore National Laboratory*, 2002.

[40] W. E. Full, R. Ehrlich, and J. E. Klovan, "EXTENDED QMODEL—objective definition of external endmembers in the analysis of mixtures," *Mathematical Geology*, vol. 13, no. 4, pp. 331–344, 1981.

[41] Y. Zhou and S. Suri, "Algorithms for a minimum volume enclosing simplex in three dimensions," *SIAM Journal on Computing*, vol. 31, no. 5, pp. 1339–1357, 2002.

[42] S. Friedberg, A. Insel, and L. Spence, *Linear Algebra*, 4th ed. Prentice Hall, Upper Saddle River, NJ, 2003.

[43] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu, and M. Zhu, "A practical algorithm for topic modeling with provable guarantees," *arXiv preprint arXiv:1212.4777*, 2012.

[44] B. A. Frigyik, A. Kapila, and M. R. Gupta, "Introduction to the Dirichlet distribution and related processes," *Tech. Rep., Department of Electrical Engineering, University of Washington, Seattle*, 2010. [Online]. Available: http://www.semanticsearchart.com/downloads/UWEETR-2010-0006.pdf

[45] J. M. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, 2012.

[46] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE IGARSS*, Barcelona, Spain, July 23-28, 2007, pp. 4033–4036.

[47] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Multivariate Distributions, Models and Applications*, 1st ed. New York: John Wiley & Sons, 2002.

[48] R. Clark, G. Swayze, R. Wise, E. Livo, T. Hoefen, R. Kokaly, and S. Sutley, "USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231," 2007. [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

[49] R. Heylen, B. Dzevdet, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4112–4122, Nov. 2011.

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

[51] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Two effective and computationally efficient pure-pixel based

algorithms for hyperspectral endmember extraction," in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 22-27, 2011, pp. 1369–1372.

[52] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, 2012.

[53] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, 2014.

[54] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems," *Journal of Geophysical Research*, vol. 108, no. 12, pp. 5–44, Dec. 2003.

[55] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summ. 4th Annu. JPL Airborne Geosci. Workshop*, vol. 2, 1992, pp. 47–49.

[56] G. Swayze, "The hydrothermal and structural history of the Cuprite Mining District, southwestern Nevada: An integrated geological and geophysical approach," Ph.D. dissertation, University of Colorado, Boulder, 1997.

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

[58] T. M. Apostol, *Mathematical Analysis*, 2nd ed.   Addison Wesley Publishing Company, 1974.