A FAST HYPERPLANE-BASED MVES ALGORITHM FOR HYPERSPECTRAL UNMIXING

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

*Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.
[†]Sunplus Technology Co., Ltd., National Science Park, Hsinchu, Taiwan, R.O.C.
E-mail: chiahsiang.steven.lin@gmail.com;cychi@ee.nthu.edu.tw;s101064505@m101.nthu.edu.tw;
chantsunghan@gmail.com

ABSTRACT

Hyperspectral unmixing (HU) is an essential signal processing procedure for blindly extracting the hidden spectral signatures of materials (or endmembers) from observed hyperspectral imaging data. Craig's criterion, stating that the vertices of the minimum volume enclosing simplex (MVES) of the data cloud yield high-fidelity endmember estimates, has been widely used for designing endmember extraction algorithms (EEAs) especially in the scenario of no pure pixels. However, most Craig-criterion-based EEAs generally suffer from high computational complexity due to heavy simplex volume computations, and performance sensitivity to random initialization, etc. In this work, based on the idea that Craig's simplex with N vertices can be defined by N associated hyperplanes, we develop a fast and reproducible EEA by identifying these hyperplanes from $N(N-1)$ data pixels extracted via simple and effective linear algebraic formulations, together with endmember identifiability analysis. Some Monte Carlo simulations are provided to demonstrate the superior efficacy of the proposed EEA over state-of-the-art Craig-criterion-based EEAs in both computational efficiency and estimation accuracy.

Index Terms— Hyperspectral unmixing, Craig's criterion, minimum volume enclosing simplex (MVES), hyperplane

1. INTRODUCTION

Hyperspectral remote sensing (HRS) is a crucial technology of imaging spectroscopy with numerous applications, such as planetary exploration, mineral identification, and military surveillance [1, 2]. The observed pixels in the hyperspectral imaging data are usually spectral mixtures of multiple substances [3] owing to limited spatial resolution of the hyperspectral sensor used. Hyperspectral unmixing (HU) [3, 4], an essential signal processing procedure for extracting individual spectral signatures of the underlying materials (or endmembers) from the measured spectral mixtures, is therefore of paramount importance in HRS.

Many existing endmember extraction algorithms (EEAs) assume the existence of pure pixels (i.e., the pixels that are solely contributed by a single endmember) [4]. Nevertheless, such pure pixel assumption (PPA) may be seriously infringed in practical applications like retinal analysis in the ophthalmology [5]. Another widely known criterion without requiring the PPA was proposed by Craig [6], stating that the vertices of the minimum-volume data-enclosing simplex are high-fidelity endmember estimates. Many EEAs based on this criterion have been proposed in the last two

decades, e.g., minimum volume constrained nonnegative matrix factorization (MVC-NMF) [7], minimum volume simplex analysis (MVSA) [8], minimum-volume enclosing simplex (MVES) [9], and simplex identification via split augmented Lagrangian (SISAL) [10], etc., but their performance and computational efficiency may be limited due to lots of complicated simplex volume calculations, sensitivity to initialization, and lack of rigorous performance analysis.

This paper proposes a fast Craig-criterion-based EEA based on the idea that Craig's simplex with N vertices can be characterized by N hyperplanes. Each hyperplane parameterized by its normal vector and a constant can be efficiently estimated from $N-1$ pixels in the data set via simple and effective linear algebraic formulations without involving any simplex volume computations. The resulting EEA, referred to as hyperplane-based Craig-simplex-identification (HyperCSI) algorithm, yields reproducible, non-negative, and, most importantly, high-fidelity endmember estimates without requiring the PPA. We also present an endmember identifiability analysis for HyperCSI algorithm. Some simulations are provided to demonstrate its superior efficacy over state-of-the-art Craig-criterion-based EEAs in both endmember estimation accuracy and computational efficiency.

Notation: conv $\mathcal{A}$ and aff $\mathcal{A}$ denote the convex hull and affine hull of a set $\mathcal{A}$, respectively [11]. $\mathbb{R}$ ($\mathbb{R}^N$, $\mathbb{R}^{M \times N}$) is the set of real numbers (N-vectors, $M \times N$ matrices). $\mathbb{R}_+^N$ ($\mathbb{R}_+^{M \times N}$) is the set of non-negative real N-vectors ($M \times N$ matrices). The set $\mathcal{I}_N = \{1, 2, \ldots, N\}$. $\mathbf{X}^\dagger$ is the pseudo-inverse of a matrix $\mathbf{X}$. $\mathbf{1}_N$ and $\mathbf{0}_N$ are all-one and all-zero N-vectors, respectively. $\mathbf{I}_N$ 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. $\mathbf{q}_i(\mathbf{X})$ denotes the ith principal eigenvector of a matrix $\mathbf{X}$ with $\|\mathbf{q}_i(\mathbf{X})\| = 1$.

2. SIGNAL MODEL AND PROBLEM STATEMENT

Consider a given hyperspectral imaging data of L pixels that consists of N distinct substances (endmembers), each characterized by a spectral signature vector $\mathbf{a}_i \in \mathbb{R}_+^M$ (where M is the number of spectral bands). Then each pixel $\mathbf{x}[n] \in \mathbb{R}^M$ in the data set can be represented as [1, 3, 4]

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

where $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_N] \in \mathbb{R}^{M \times N}$ is the spectral signature matrix and $\mathbf{s}[n] = [s_1[n] \cdots s_N[n]]^T \in \mathbb{R}^N$ is the abundance vector. In this work, we assume that $N \geq 3$ is known a priori as it can be estimated using model-order selection methods, such as virtual dimensionality (VD) [12] and hyperspectral signal subspace identification by minimum error (HySiMe) [13].

Hyperspectral unmixing is to blindly extract the N unknown endmembers (i.e., $\mathbf{a}_1, \ldots, \mathbf{a}_N$) from the observed spectral data

$\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}$. Some standard assumptions pertaining to the linear mixing model (1) are as follows [1, 3, 4]:

(A1) $s_i[n] \geq 0$, for all $i \in \mathcal{I}_N$ and $n \in \mathcal{I}_L$.

(A2) $\sum_{i=1}^N s_i[n] = 1$, for all $n \in \mathcal{I}_L$.

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

Under the above assumptions, the pixel $\mathbf{x}[n]$ in the original image can be equivalently represented in a dimension-reduced (DR) space via affine set fitting [14] as follows:

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

where

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

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

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

in which $\mathbf{U} = [\,\mathbf{x}[1] - \mathbf{d}, \ldots, \mathbf{x}[L] - \mathbf{d}\,] \in \mathbb{R}^{M \times L}$ (mean removed data matrix), $\mathbf{C}$ is semi-unitary (i.e., $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{N-1}$), and $\mathbf{d}$ corresponds to the origin $\mathbf{0}_{N-1}$ in the DR space $\mathbb{R}^{N-1}$ (by (2)).

From (2) and (A1)-(A2), it can be seen that

$$\mathcal{X} \triangleq \{\, \tilde{\mathbf{x}}[1], \ldots, \tilde{\mathbf{x}}[L]\,\} \subseteq \mathsf{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}, \tag{6}$$

i.e., the true endmembers' simplex $\mathsf{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subseteq \mathbb{R}^{N-1}$ itself is a data-enclosing simplex (in the noiseless scenario). By Craig's criterion, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$ are estimated by solving the following volume minimization problem [9]:

$$\begin{aligned} \min_{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N} \quad & V(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N) \\ \text{s.t.} \quad & \tilde{\mathbf{x}}[n] \in \mathsf{conv}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}, \ \forall n, \end{aligned} \tag{7}$$

where $V(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)$ denotes the volume of the simplex $\mathsf{conv}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\} \subseteq \mathbb{R}^{N-1}$. Under some mild conditions on data purity level, the optimal solution of the problem (7) can perfectly yield the true endmembers in the absence of pure pixels [15, 16].

3. HYPERPLANE-BASED CSI ALGORITHM

In this section, without involving any simplex volume computations, we propose a computationally efficient and performance effective algorithm based on the idea stated in the following proposition:

Proposition 1 *If* $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subseteq \mathbb{R}^{N-1}$ *is affinely independent (i.e.,* $\{\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_N, \ldots, \boldsymbol{\alpha}_{N-1} - \boldsymbol{\alpha}_N\}$ *is linearly independent), then the simplex* $\mathcal{T} = \mathsf{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subseteq \mathbb{R}^{N-1}$ *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 $\{\mathcal{H}_1, \ldots, \mathcal{H}_N\}$. Note that hyperplane $\mathcal{H}_i$ can be parameterized by a normal vector $\mathbf{b}_i \in \mathbb{R}^{N-1}$ and a constant $h_i \in \mathbb{R}$ as [11]

$$\mathcal{H}_i(\mathbf{b}_i, h_i) = \{\,\mathbf{x} \in \mathbb{R}^{N-1} \mid \mathbf{b}_i^T\mathbf{x} = h_i\,\}. \tag{8}$$

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 (8) that $\mathbf{b}_j^T\boldsymbol{\alpha}_i = h_j$ for all $j \neq i$, i.e.,

$$\mathbf{B}_{-i}\boldsymbol{\alpha}_i = \mathbf{h}_{-i}, \tag{9}$$

where

$$\mathbf{B}_{-i} \triangleq [\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \ldots, \mathbf{b}_N]^T \in \mathbb{R}^{(N-1) \times (N-1)} \tag{10}$$

$$\mathbf{h}_{-i} \triangleq [h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_N]^T \in \mathbb{R}^{N-1}. \tag{11}$$

As $\mathcal{T}$ is a simplex in $\mathbb{R}^{N-1}$, $\mathbf{B}_{-i}$ must be of full rank and hence invertible [11]. Hence, we have from (9)

$$\boldsymbol{\alpha}_i = \mathbf{B}_{-i}^{-1}\mathbf{h}_{-i}, \ \forall i \in \mathcal{I}_N, \tag{12}$$

implying that the simplex $\mathcal{T}$ can be reconstructed. ∎

Moreover, the outward-pointing normal vector $\mathbf{b}_i$ (cf. Figure 1) of hyperplane $\mathcal{H}_i$ given by (8) has a closed-form expression [17]:

$$\begin{aligned} \mathbf{b}_i &\equiv \boldsymbol{v}_i(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) \tag{13} \\ &\triangleq \left(\mathbf{I}_{N-1} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right) \cdot (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_i), \text{ for any } j \neq i, \end{aligned}$$

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 (13) for obtaining the normal vector $\mathbf{b}_i$ of $\mathcal{H}_i$, one can show that $\mathbf{b}_i$ can also be alternatively obtained from any given affinely independent set $\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\} \subseteq \mathcal{H}_i$ as follows:

$$\mathbf{b}_i = \boldsymbol{v}_i(\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{i-1}^{(i)}, \mathbf{0}_{N-1}, \mathbf{p}_i^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}), \tag{14}$$

where $\boldsymbol{v}_i(\cdot)$ is defined in (13). The proof of (14) is omitted due to space limitation.

It can be inferred from (A3) that the set of DR endmembers $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ is affinely independent. By Proposition 1, problem (7) can be decoupled into N subproblems of hyperplane estimation or equivalently, estimation of the parameter vectors $(\mathbf{b}_i, h_i)$, $i \in \mathcal{I}_N$ (cf. (8)). In the following subsections, let us present how to estimate $\mathbf{b}_i$ and h_i from the DR data set $\mathcal{X}$, respectively.

3.1. Normal Vector Estimation

Based on (14), the idea of determining the normal vector of $\mathcal{H}_i$ is to find $N-1$ affinely independent points

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

that are as close to $\mathcal{H}_i$ as possible. Another important observation from (6) is given in the following fact:

(F1) All the pixels in $\mathcal{X}$ lie on the same side of $\mathcal{H}_i$ (cf. (6)), implying that the pixel $\mathbf{p} \in \mathcal{X}$ closest to $\mathcal{H}_i$ is exactly the one with maximum value of $\mathbf{b}_i^T\mathbf{p}$.

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 can be obtained using the reliable and reproducible TRIP algorithm [18]. So $\tilde{\boldsymbol{\alpha}}_i$ can be viewed as the pixel in $\mathcal{X}$ "closest" to $\boldsymbol{\alpha}_i$ (cf. Figure 1). Let $\tilde{\mathbf{b}}_i$ be the outward-pointing normal vector of hyperplane $\widetilde{\mathcal{H}}_i \triangleq \mathsf{aff}(\, \{\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N\} \setminus \{\tilde{\boldsymbol{\alpha}}_i\}\,)$, i.e.,

$$\tilde{\mathbf{b}}_i \triangleq \boldsymbol{v}_i(\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N). \ (\text{cf. (13)}) \tag{15}$$

Considering (F1) and that $\mathcal{P}_i$ must contain $N-1$ distinct pixels, we search for the desired affinely independent set $\mathcal{P}_i$ by:

$$\mathbf{p}_k^{(i)} \in \ \arg\max \ \{\tilde{\mathbf{b}}_i^T\mathbf{p} \mid \mathbf{p} \in \mathcal{X} \cap \mathcal{R}_k^{(i)}\}, \ \forall k \in \mathcal{I}_{N-1}, \tag{16}$$

where

$$\mathcal{R}_k^{(i)} \triangleq \begin{cases} \mathcal{B}(\tilde{\boldsymbol{\alpha}}_k, r), \text{ if } k < i, \\ \mathcal{B}(\tilde{\boldsymbol{\alpha}}_{k+1}, r), \text{ if } k \geq i, \end{cases} \tag{17}$$

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 $\mathcal{X} \cap \mathcal{R}_k^{(i)} \neq \emptyset$ (as
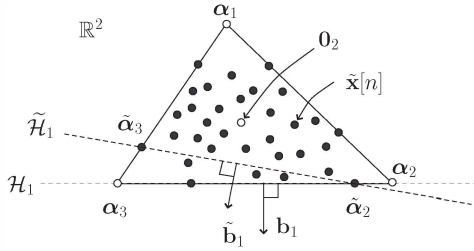
Fig. 1. An illustration of hyperplanes and DR data in $\mathbb{R}^2$ for the case of $N = 3$, where $\tilde{\boldsymbol{\alpha}}_3$ is a purest pixel in $\mathcal{X}$ (a purest pixel $\tilde{\boldsymbol{\alpha}}_i$ can be considered as the pixel closest to $\boldsymbol{\alpha}_i$) but not very close to hyperplane $\mathcal{H}_1 = \mathsf{aff}\{\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3\}$, leading to nontrivial orientation difference between $\tilde{\mathbf{b}}_1$ and $\mathbf{b}_1$.

it contains either $\tilde{\boldsymbol{\alpha}}_k$ or $\tilde{\boldsymbol{\alpha}}_{k+1}$, cf. (17)), i.e., problem (16) is feasible. Then we obtain the estimated normal vector associated with $\mathcal{H}_i$ as

$$\tilde{\mathbf{b}}_i = \boldsymbol{v}_i(\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{i-1}^{(i)}, \mathbf{0}_{N-1}, \mathbf{p}_i^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}). \ \ (\text{cf. (14)}) \ \ (18)$$

In addition to assumptions (A1)-(A3), with one more assumption that is extensively used to characterize the behavior of the abundance vectors in the HRS context [19, 20]:

(A4) the abundance vectors $\{\mathbf{s}[n]\} \subseteq \mathbb{R}^N$ are independent and identically distributed (i.i.d.) following the Dirichlet distribution [21] with parameter vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_N]^T \succ \mathbf{0}_N$,

the obtained $\mathcal{P}_i$ by (16) can be proved to be affinely independent as stated in the following theorem (with proof given in Appendix):

Theorem 1 *Assume (A1)-(A4) hold true. Let $\mathbf{p}_k^{(i)} \in \mathcal{P}_i$ be a solution to (16) with $\mathcal{R}_k^{(i)}$ defined in (17), 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).*

Note that the orientation difference between $\tilde{\mathbf{b}}_i$ and the true $\mathbf{b}_i$ may not be small (cf. Figure 1). Hence, $\tilde{\mathbf{b}}_i$ itself may not be a good estimate for $\mathbf{b}_i$. Nevertheless, it can be shown that the orientation difference between $\hat{\mathbf{b}}_i$ and $\mathbf{b}_i$ tends to be very small for large L, even in the absence of pure pixels (as stated in Remark 1 below).

3.2. Hyperplane Estimation and Performance Analysis

With the estimated normal vector $\hat{\mathbf{b}}_i$ (cf. (18)), as the hyperplanes associated with the minimum-volume data-enclosing simplex must be externally tangent to the data cloud $\mathcal{X}$, they can be determined as $\mathcal{H}_i(\hat{\mathbf{b}}_i, \hat{h}_i), \forall i \in \mathcal{I}_N$, where $\hat{h}_i$ is obtained by solving

$$\hat{h}_i = \max \ \{\ \hat{\mathbf{b}}_i^T \mathbf{p} \mid \mathbf{p} \in \mathcal{X} \ \}. \quad (19)$$

Considering the volume expansion due to noise effect [22, 23], the estimated hyperplanes need to be properly shifted closer to the origin, so instead, $\mathcal{H}_i(\hat{\mathbf{b}}_i, \hat{h}_i/c), i \in \mathcal{I}_N$, are the desired hyperplane estimates for some $c \geq 1$. Therefore, the corresponding DR endmember estimates are obtained by (cf. (12))

$$\hat{\boldsymbol{\alpha}}_i = \widehat{\mathbf{B}}_{-i}^{-1} \cdot \hat{\mathbf{h}}_{-i}/c, \quad \forall i \in \mathcal{I}_N, \quad (20)$$

where $\widehat{\mathbf{B}}_{-i}$ and $\hat{\mathbf{h}}_{-i}$ are given by (10) and (11) with $\mathbf{b}_j$ and h_j replaced by $\hat{\mathbf{b}}_j$ and $\hat{h}_j, \forall j \neq i$, respectively. It is necessary to choose c such that the associated endmember estimates in the original space

$$\hat{\mathbf{a}}_i = \mathbf{C} \hat{\boldsymbol{\alpha}}_i + \mathbf{d} \succeq \mathbf{0}_M, \forall i \in \mathcal{I}_N. \ \text{(cf. (5) and (A3))} \quad (21)$$

By (20) and (21), it is required that $c \geq c'$ where

$$c' \triangleq \min_{c'' \geq 1}\{c'' \mid \mathbf{C}(\widehat{\mathbf{B}}_{-i}^{-1} \cdot \hat{\mathbf{h}}_{-i}) + c'' \cdot \mathbf{d} \succeq \mathbf{0}_M, \forall i\} \quad (22)$$

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

$$c' = \max\left\{1, \max\{-v_{ij}/d_j \mid i \in \mathcal{I}_N, \ j \in \mathcal{I}_M\}\right\}, \quad (23)$$

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 need to 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 signal-to-noise ratio (SNR) greater than 20 dB; typically the value of SNR in real hyperspectral data is much higher than 20 dB, e.g., AVIRIS [24]. 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 $\mathsf{conv}\{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_N\}$ to the boundary of the non-negative orthant $\mathbb{R}_+^M$. On the other hand, we empirically observed that typical endmembers in the U.S. geological survey (USGS) library [25] are close to the boundary of $\mathbb{R}_+^M$. That is to say, a reasonable choice of $\eta \in (0, 1]$ should be large (i.e., close to 1), accounting for the reason why the preset value of $\eta = 0.9$ can always yield good performance. The resulting HyperCSI algorithm is summarized in Table 1.

Table 1. 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 TRIP algorithm [18].
Step 3.	Obtain $\tilde{\mathbf{b}}_i$ using (13), $\forall i$, and $\mathcal{R}_k^{(i)}$ using (17), $\forall i, k$.
Step 4.	Obtain $(\mathcal{P}_i, \hat{\mathbf{b}}_i, \hat{h}_i)$ by (16), (18), and (19), $\forall i \in \mathcal{I}_N$.
Step 5.	Obtain c' by (23), and set $c = c'/\eta$.
Step 6.	Calculate $\hat{\boldsymbol{\alpha}}_i$ by (20) and $\hat{\mathbf{a}}_i = \mathbf{C} \hat{\boldsymbol{\alpha}}_i + \mathbf{d}$ by (21), $\forall i$.
Output	The estimated endmembers $\{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_N\}$.

Asymptotic identifiability of the proposed HyperCSI algorithm can be guaranteed as stated in the following theorem:

Theorem 2 *Under (A1)-(A4), the noiseless assumption and $L \to \infty$, the simplex identified by HyperCSI algorithm with $c = 1$ is exactly the Craig's minimum-volume simplex (i.e., solution of (7)) and the true endmembers' simplex $\mathsf{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ in the DR space w.p.1.*

The proof is omitted due to space limit. Instead, the philosophies behind the proof of Theorem 2 are given in the following two remarks:

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). That is to say, $\hat{\mathbf{b}}_i$ can be uniquely obtained by (18), and its orientation approaches to that of $\mathbf{b}_i$ w.p.1.

Remark 2 Remark 1 together with (6) implies that $\hat{h}_i$ is upper bounded by h_i w.p.1 (assuming 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 $h_i = \hat{h}_i/c$ 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. (20)) as $L \to \infty$ in the absence of noise. Actually, with a moderate L and finite SNR, the proposed HyperCSI algorithm can yield high-fidelity endmember estimates as demonstrated in the simulation results below.

4. SIMULATION RESULTS

Six endmembers (i.e., Jarsoite, Pyrope, Dumortierite, Buddingtonite, Muscovite, and Goethite) with $M = 224$ spectral bands randomly selected from the USGS library [25] are used to generate $L = 10000$ synthetic hyperspectral data $\mathbf{x}[n]$, where the abundance vectors $\mathbf{s}[n]$ are i.i.d. generated according to Dirichlet distribution with parameter $\boldsymbol{\gamma} = \mathbf{1}_N/N$ (automatically enforcing (A1)-(A2)) for various values of SNR (Gaussian noise added) and different data purity levels $\rho = \max\{\|\mathbf{s}[n]\|, n \in \mathcal{I}_L\}$ [9, 15, 23]. The average root-mean-square (RMS) spectral angle ϕ_{en} between the true endmembers $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ and their estimates $\{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_N\}$ [9, 26] over 100 independent runs is used as the performance measure for comparison of the proposed HyperCSI algorithm and four benchmarked Craig-criterion-based EEAs, including MVC-NMF [7], MVSA [8], MVES [9], and SISAL [10]. It should be mentioned that the performances of these four EEAs are dependent on their respective regularization parameters, and we have tried our best to select these parameters so as to yield their best performances.

The simulation results of average RMS spectral angle ϕ_{en} and average computation time T per realization are shown in Table 2, where bold-face numbers indicate the best performance (i.e., the smallest ϕ_{en} or T) for a specific scenario of ρ and SNR. From this table, it can be seen that HyperCSI algorithm significantly outperforms all the other EEAs in terms of ϕ_{en} and T for almost all the cases, especially for lower value of SNR or lower value of ρ, while SISAL outperforms the other three EEAs for SNR > 20dB. These results also indicate that $L = 10000$ (typically several ten thousands in HRS applications) is large enough for the proposed algorithm to achieve the asymptotic performance as stated in Theorem 2.

Table 2. Performance comparison of the proposed HyperCSI algorithm and four state-of-the-art Craig-criterion-based EEAs.

Methods	ρ	ϕ_{en} (degrees)					T (seconds)
		SNR (dB)					
		20	25	30	35	40	
MVC-NMF	0.8	2.87	2.38	1.50	1.24	1.18	
	0.9	2.98	1.87	0.93	0.54	0.44	347.77
	1	3.26	1.90	1.01	0.52	**0.21**	
MVSA	0.8	11.05	6.23	3.41	1.87	1.03	
	0.9	11.58	6.46	3.48	1.90	1.05	3.18
	1	11.65	6.54	3.54	1.93	1.06	
MVES	0.8	10.66	6.06	3.39	1.91	1.16	
	0.9	10.17	6.06	3.48	1.97	1.12	27.97
	1	9.95	5.96	3.55	2.19	1.30	
SISAL	0.8	4.01	2.31	1.30	0.72	**0.39**	
	0.9	4.19	2.43	1.36	0.74	0.40	0.69
	1	4.49	2.59	1.45	0.79	0.42	
HyperCSI	0.8	**1.88**	**1.28**	**0.91**	**0.61**	**0.39**	
	0.9	**1.34**	**0.90**	**0.61**	**0.44**	**0.31**	**0.43**
	1	**1.15**	**0.78**	**0.54**	**0.37**	0.26	

5. CONCLUSIONS

We have presented a new fast Craig-criterion-based EEA, called HyperCSI algorithm, given in Table 1, based on the convex geometry concept—hyperplane. It has several remarkable characteristics:

- It never requires the presence of pure pixels in the data.
- It is reproducible without involving random initialization.
- It estimates Craig's minimum-volume simplex by finding only $N(N-1)$ pixels (regardless of L, cf. (16)) without involving any simplex volume computation, accounting for its high computational efficiency.
- The estimated endmembers are guaranteed non-negative, and the identified simplex was proven to be both Craig's simplex and true endmembers' simplex as $L \to \infty$ for the noiseless case w.p.1.

Simulation results also demonstrated its superior efficacy over some state-of-the-art algorithms in both solution accuracy and computational efficiency.

6. APPENDIX: PROOF OF THEOREM 1

For a fixed $i \in \mathcal{I}_N$, one can see from (17) 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 (16) 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

$$\mathbf{p}_k \neq \mathbf{p}_\ell, \text{ for all } 1 \leq k < \ell \leq N-1. \quad (24)$$

Then, as $\mathbf{p}_k \in \mathcal{X}$, $\forall k \in \mathcal{I}_{N-1}$, we have from (A4) and (24) that there exist i.i.d Dirichlet distributed random vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\}$ such that (cf. (2))

$$\mathbf{p}_k = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_N] \mathbf{s}_k, \quad \text{for all } k \in \mathcal{I}_{N-1}. \quad (25)$$

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

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

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

$$\mathbf{p}_1 = \theta_2 \cdot \mathbf{p}_2 + \cdots + \theta_{N-1} \cdot \mathbf{p}_{N-1}, \quad (26)$$

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

$$\theta_2 + \cdots + \theta_{N-1} = 1. \quad (27)$$

By substituting (25) into (26), we have

$$[\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_N] \mathbf{s}_1 = \sum_{m=2}^{N-1} [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_N] (\theta_m \cdot \mathbf{s}_m). \quad (28)$$

Then, from the fact that $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ is affinely independent (cf. (A3)) and the fact that $\mathbf{1}_N^T(\sum_{m=2}^{N-1} (\theta_m \cdot \mathbf{s}_m)) = 1$ (by (27) and the fact that $\mathbf{1}_N^T \mathbf{s}_k = 1$, $\forall k$), (28) implies

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

which together with (27) further implies that E2 is true. Thus we have proved that E1 implies E2, and hence

$$\Pr\{E1\} \leq \Pr\{E2\}. \quad (29)$$

As Dirichlet distribution is a continuous multivariate distribution [27] 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 [21]

$$\Pr\{\mathbf{s} \in \mathcal{A}\} = 0, \text{ if } P < N-1. \quad (30)$$

Moreover, as $\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\}$ are i.i.d. random vectors and the affine hull $\mathsf{aff}\{\{\mathbf{s}_1, \ldots, \mathbf{s}_{N-1}\} \setminus \{\mathbf{s}_k\}\}$ must have affine dimension $P < N-1$, we have from (30) that

$$\Pr\{E3^{(k)}\} = 0, \text{ for all } k \in \mathcal{I}_{N-1}. \quad (31)$$

Then we have the following inferences:

$$0 \leq \Pr\{E1\} \leq \Pr\{E2\} \quad \text{(by (29))}$$
$$= \Pr\{\cup_{k=1}^{N-1} E3^{(k)}\} \quad \text{(by the definitions of E2 and E3}^{(k)})$$
$$\leq \sum_{k=1}^{N-1} \Pr\{E3^{(k)}\} = 0, \quad \text{(by the union bound and (31))}$$

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

7. REFERENCES

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

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

[3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, 2012.

[4] 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.

[5] 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.

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

[7] 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.

[8] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, Boston, MA, July 7-11, 2008, pp. 250–253.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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," *arXiv preprint arXiv:1406.5273*, 2014.

[16] C.-H. Lin, A. Ambikapathi, W.-C. Li, and C.-Y. Chi, "On the endmember identifiability of Craig's criterion for hyperspectral unmixing: A statistical analysis for three-source case," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 26-31, 2013, pp. 2139–2143.

[17] G. Strang, *Linear Algebra and Its Applications*, 4th ed. CA: Thomson, 2006.

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

[19] 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.

[20] 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.

[21] 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

[22] 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.

[23] 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.

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

[25] 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

[26] 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.

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