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: \( \text{conv} \ A \) and \( \text{aff} \ A \) denote the convex hull and affine hull of a set \( A \), respectively [11]. \( \mathbb{R}^{N, M} \) is the set of real numbers \((N\text{-vectors, } M \times N \text{ matrices})\). \( \mathbb{R}_{\geq 0}^{M \times N} \) is the set of non-negative real \( N\)-vectors \((M \times N \text{ matrices})\). The set \( \mathcal{I}_N = \{1, 2, \ldots, N\} \) is the pseudo-inverse of a matrix \( X \). \( \mathbf{1}_N \) and \( \mathbf{0}_N \) are all-one and all-zero \( N\)-vectors, respectively. \( \mathbf{1}_N \) is the \( N \times N \) identity matrix.

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 \( x[n] \in \mathbb{R}^M \) in the data set can be represented as [1,3,4]

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

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
\{x[1], \ldots, 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 \(A \in \mathbb{R}^{L \times N}\) is of full column rank.

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

\[
x[n] = C^T (x[n] - d) = \sum_{i=1}^{N} s_i[n] \alpha_i \in \mathbb{R}^{N-1},
\]

where

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

\[
C = [ q_1(UU^T), \ldots, q_{N-1}(UU^T) ] \in \mathbb{R}^{N \times (N-1)}
\]

\[
\alpha = C^T (\alpha - d) \in \mathbb{R}^{N-1} \text{ (endmembers in the DR space)}
\]

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

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

\[
\mathcal{X} \triangleq \{ x[1], \ldots, x[L] \} \subseteq \text{conv} \{\alpha_1, \ldots, \alpha_N\},
\]

i.e., the true endmembers’ simplex \(\text{conv} \{\alpha_1, \ldots, \alpha_N\} \subseteq \mathbb{R}^{N-1}\) itself is a data-enclosing simplex (in the noiseless scenario). By Craig’s criterion, \(\alpha_1, \ldots, \alpha_N\) are estimated by solving the following volume minimization problem [9]:

\[
\begin{aligned}
\min_{\beta_1, \ldots, \beta_N} & \quad V(\beta_1, \ldots, \beta_N) \\
\text{s.t.} & \quad \hat{x}[n] \in \text{conv} \{\beta_1, \ldots, \beta_N\}, \quad \forall \, n,
\end{aligned}
\]

where \(V(\beta_1, \ldots, \beta_N)\) denotes the volume of the simplex \(\text{conv} \{\beta_1, \ldots, \beta_N\} \subseteq \mathbb{R}^{N-1}\). 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 \(\{\alpha_1, \ldots, \alpha_N\} \subseteq \mathbb{R}^{N-1}\) is affinely independent (i.e., \(\{\alpha_1 - \alpha_N, \ldots, \alpha_{N-1} - \alpha_N\}\) is linearly independent), then the simplex \(\mathcal{T} = \text{conv} \{\alpha_1, \ldots, \alpha_N\} \subseteq \mathbb{R}^{N-1}\) can be reconstructed from the associated \(N\) hyperplanes \(\{\mathcal{H}_1, \ldots, \mathcal{H}_N\}\), that tightly enclose \(\mathcal{T}\), where \(\mathcal{H}_i \triangleq \text{aff} \{\{\alpha_1, \ldots, \alpha_N\} \setminus \{\alpha_i\}\}\), i.e.,

\[
\mathcal{B}_i \triangleq v_i(\alpha_1, \ldots, \alpha_N). \quad \text{(cf. (13))}
\]

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

\[
p^{(i)}_k \in \arg \max \{ \mathcal{B}_i^T p \mid p \in \mathcal{X} \cap \mathcal{R}_k^{(i)} \}, \quad \forall \, k \in \mathcal{I}_{N-1},
\]

where

\[
\mathcal{R}_k^{(i)} \triangleq \left\{ \mathcal{B}(\alpha_k, r), \text{if } k < i, \right. \\
\left. \mathcal{B}(\alpha_{k+1}, r), \text{if } k \geq i, \right\}
\]

As \(\alpha_i \in \text{aff} \{\{\alpha_1, \ldots, \alpha_N\} \setminus \{\alpha_i\}\}\), for all \(j \neq i\), we have from (8) that \(B_j^T \alpha_i = h_j\) for all \(j \neq i\), i.e.,

\[
B_j^T \alpha_i = h_j, \quad \forall \, j \neq i,
\]

where

\[
B_j = [b_{1j}, \ldots, b_{ij-1}, b_{ij+1}, \ldots, b_{Lj}]^T \in \mathbb{R}^{(N-1) \times (N-1)}
\]

\[
h_j = [h_{1j}, \ldots, h_{ij-1}, h_{ij+1}, \ldots, h_{Lj}]^T \in \mathbb{R}^{N-1}.
\]
it contains either \(\hat{\alpha}_k\) or \(\hat{\alpha}_{k+1}\), cf. (17), i.e., problem (16) is feasible. Then we obtain the estimated normal vector associated with \(H_i\) as

\[
b_i = v_o(p_{i1}^{(i)}, \ldots, p_{iN-1}^{(i)}, 0_{N-1}, p_{i1}^{(i)}, \ldots, p_{iN-1}^{(i)}).
\]

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 \(\{s[n]\} \subseteq R^N\) are independent and identically distributed (i.i.d.) following the Dirichlet distribution [21] with parameter vector \(\gamma = [\gamma_1, \ldots, \gamma_N]^T \succ 0_N\), the obtained \(P_i\) 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 \(p_i^{(i)} \in P_i\) be a solution to (16) with \(R_i^{(i)}\) defined in (17), for all \(i \in I_N\) and \(k \in I_{N-1}\). Then, the set \(P_i\) is affinely independent with probability 1 (w.p.1).

Note that the orientation difference between \(b_i\) and the true \(b_i\) may not be small (cf. Figure 1). Hence, \(b_i\) itself may not be a good estimate for \(b_i\). Nevertheless, it can be shown that the orientation difference between \(b_i\) and \(b_i\) 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 \(b_i\), (18), as the hyperplanes associated with the minimum-volume data-enclosing simplex must be externally tangent to the data cloud \(X\), they can be determined as \(H_i(b_i, h_i), \forall i \in I_N\), where \(h_i\) is obtained by solving

\[
h_i = \max \{h_i^T p | p \in X\}\]  

(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, \(H_i(b_i, h_i/c), \forall i \in I_N\), are the desired hyperplane estimates for some \(c \geq 1\). Therefore, the corresponding DR endmember estimates are obtained by (cf. (12))

\[
\hat{\alpha}_i = \tilde{B}_i^{-1} \cdot h_{i/c}, \forall i \in I_N\]  

(20)

where \(\tilde{B}_i\) and \(h_{i/c}\) are given by (10) and (11) with \(b_i\) replaced by \(\tilde{B}_i\) and \(h_{i/c}\). By (21), it is required that \(c \geq c'\) where

\[
c' \triangleq \min_{c'' \geq 1} \{c'' | C \cdot (\tilde{B}_i^{-1} \cdot h_{i/c}) + c'' \cdot d \geq 0_M, \forall i\}\]  

(21)

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

\[
c' = \max \{1, \max_{i \in I_N} \{-d_j/J | j \in I_M\}\}\]  

(22)

where \(v_j\) is the \(j\)th component of \(C \cdot (\tilde{B}_i^{-1} \cdot h_{i/c}) + d\) and \(d_j\) is the \(j\)th component of \(d\).

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 \(\eta\) (or the smaller the value of \(c'\)), the farther the estimated hyperplanes from the origin \(0_{N-1}\), or the closer the estimated endmembers’ simplex \(\text{conv}\{\hat{\alpha}_1, \ldots, \hat{\alpha}_N\}\) to the boundary of the non-negative orthant \(R^M_+\). 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 \(R^M_+\). 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>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1.</td>
<td>Calculate (C, d) using (3)-(4), and obtain the DR data (X = [x[1], \ldots, x[L]]) using (2).</td>
</tr>
<tr>
<td>Step 2.</td>
<td>Obtain (\hat{\alpha}_1, \ldots, \hat{\alpha}_N) using TRIP algorithm [18].</td>
</tr>
<tr>
<td>Step 3.</td>
<td>Obtain (\tilde{B}_i) using (13), (\forall i), and (R_i^{(i)}) using (17), (\forall i, k).</td>
</tr>
<tr>
<td>Step 4.</td>
<td>Obtain (P_i) by (16), (18), and (19), (\forall i \in I_N).</td>
</tr>
<tr>
<td>Step 5.</td>
<td>Obtain (c') by (22), and set (c = c'/\eta).</td>
</tr>
<tr>
<td>Step 6.</td>
<td>Calculate (\hat{\alpha}_i) by (20) and (\hat{\alpha}_i = C \cdot \hat{\alpha}_i + d) by (21), (\forall i).</td>
</tr>
</tbody>
</table>

**Output** The estimated endmembers \(\{\hat{\alpha}_1, \ldots, \hat{\alpha}_N\}\).

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 \(L \rightarrow \infty\), 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 \(\text{conv}\{\alpha_1, \ldots, \alpha_N\}\) 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 \(P_i\) can be shown to be arbitrarily close to \(H_i\), as the pixel number \(L \rightarrow \infty\), and they are affinely independent w.p.1 (cf. Theorem 1). That is to say, \(b_i\) can be uniquely obtained by (18), and its orientation approaches to that of \(b_i\) 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 \(|b_i| = |b_i|\)), and this upper bound can be shown to be achievable w.p.1 as \(L \rightarrow \infty\). Thus, as \(c = 1\), we have that \(h_i = h_i/c\) w.p.1.

It can be further inferred, from the above two remarks, that \(\hat{\alpha}_i\) is exactly the true \(\alpha_i\) w.p.1 (cf. (20)) as \(L \rightarrow \infty\) 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 \( x[n] \), where the abundance vectors \( s[n] \) are i.i.d. generated according to Dirichlet distribution with parameter \( \gamma = 1/N \) (automatically enforcing (A1)-(A2)) for various values of SNR (Gaussian noise added) and different data purity levels \( \rho = \max\{||s[n]||, n \in \mathcal{I}_L\} \) [9, 15, 23]. The average root-mean-square (RMS) spectral angle \( \phi_{en} \) between the true endmembers \( \{a_1, \ldots, a_N\} \) and their estimates \( \{\hat{a}_1, \ldots, \hat{a}_N\} \) [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 EAs, including MVC-NMF [7], MVSA [8], MVES [9], and SISAL [10]. It should be mentioned that the performances of these four EAs 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 \( \phi_{en} \) and average computation time \( T \) per realization are shown in Table 2, where bold-face numbers indicate the best performance (i.e., the smallest \( \phi_{en} \) or \( T \)) for a specific scenario of \( \rho \) and SNR. From this table, it can be seen that HyperCSI algorithm significantly outperforms all the other EAs in terms of \( \phi_{en} \) and \( T \) for almost all the cases, especially for lower value of SNR or lower value of \( \rho \), while SISAL outperforms the other three EAs 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 EAs.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( \rho )</th>
<th>( \phi_{en} ) (degrees)</th>
<th>SNR (dB)</th>
<th>( T ) (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>MVC-NMF</td>
<td>0.8</td>
<td>2.87</td>
<td>2.38</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>2.98</td>
<td>1.87</td>
<td>0.93</td>
</tr>
<tr>
<td>MVSA</td>
<td>0.8</td>
<td>11.05</td>
<td>6.23</td>
<td>3.44</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>11.58</td>
<td>6.46</td>
<td>3.48</td>
</tr>
<tr>
<td>MVES</td>
<td>0.8</td>
<td>10.06</td>
<td>6.06</td>
<td>3.39</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>10.17</td>
<td>6.06</td>
<td>3.48</td>
</tr>
<tr>
<td>SISAL</td>
<td>0.8</td>
<td>4.01</td>
<td>2.31</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>4.19</td>
<td>2.43</td>
<td>1.36</td>
</tr>
<tr>
<td>HyperCSI</td>
<td>0.8</td>
<td>1.88</td>
<td>1.28</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1.34</td>
<td>0.90</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>1.15</td>
<td>0.75</td>
<td>0.85</td>
<td>0.37</td>
</tr>
</tbody>
</table>

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 \( \mathcal{R}_k^i \cap \mathcal{R}_k^i = \emptyset \), \( \forall k \neq \ell \), implying that the \( N - 1 \) pixels \( p_k^i, \forall k \in \mathcal{I}_{N-1} \), identified by solving (16) must be distinct. Hence, it suffices to show that \( \mathcal{P} \) is affinely independent w.p.1 for any \( \mathcal{P} \triangleq \{p_1, \ldots, p_{N-1}\} \subseteq \mathcal{X} \) that satisfies \( p_k \neq p_i \), for all \( 1 \leq k < \ell \leq N - 1 \). (24)

Then, as \( p_i \in \mathcal{X} \), \( \forall k \in \mathcal{I}_{N-1} \), we have from (A4) and (24) that there exist i.i.d Dirichlet distributed random vectors \( \{s_1, \ldots, s_{N-1}\} \) such that (cf. (2)) \( p_k = [\alpha_1 \cdots \alpha_N] s_k \), for all \( k \in \mathcal{I}_{N-1} \). (25)

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

- \( E_1 \) The set \( \mathcal{P} \) is affinely dependent.
- \( E_2 \) The set \( \{s_1, \ldots, s_{N-1}\} \) is affinely dependent.

Then, to prove that \( P_i \) is affinely independent w.p.1, it suffices to prove \( Pr\{E_1\} = 0 \).

Next, let us show that \( E_1 \) implies \( E_2 \). Assume \( E_1 \) is true. Then \( p_k \in aff\{\mathcal{P} \setminus \{p_i\}\} \) for some \( k \in \mathcal{I}_{N-1} \). Without loss of generality, let us assume \( k = 1 \). Then,

\[
p_1 = \theta_2 p_2 + \cdots + \theta_{N-1} p_{N-1},
\]

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

\[
\theta_2 + \cdots + \theta_{N-1} = 1.
\]

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

\[
\{\alpha_1 \cdots \alpha_N\} s_1 = \sum_{m=2}^{N-1} [\alpha_1 \cdots \alpha_N] (\theta_m \cdot s_m).
\]

Then, from the fact that \( \{\alpha_1, \ldots, \alpha_N\} \) is affinely independent (cf. (A3)) and the fact that \( \sum_{m=2}^{N-1} (\theta_m \cdot s_m) = 1 \) (by (27) and the fact that \( \sum_{m=2}^{N-1} s_m = 1, \forall k \)), (28) implies

\[
s_1 = \theta_2 s_2 + \cdots + \theta_{N-1} s_{N-1},
\]

which together with (27) further implies that \( E_2 \) is true. Thus we have proved that \( E_1 \) implies \( E_2 \), and hence

\[
Pr\{E_1\} \leq Pr\{E_2\}.
\]

As Dirichlet distribution is a continuous multivariate distribution [27] for a random vector \( s \in \mathcal{R}^N \) to satisfy (A1)-(A2) with an \( (N-1) \)-dimensional domain, any given affine hull \( A \subseteq \mathcal{R}^N \) with affine dimension \( p \) must satisfy [21]

\[
Pr\{s \in A\} = 0, \quad \text{if} \quad p < N - 1.
\]

Moreover, as \( \{s_1, \ldots, s_{N-1}\} \) are i.i.d. random vectors and the affine hull \( aff\{\{s_1, \ldots, s_{N-1}\} \setminus \{s_k\}\} \) must have affine dimension \( p < N - 1 \), we have from (30) that

\[
Pr\{E_3^{(k)}\} = 0, \quad \text{for all} \quad k \in \mathcal{I}_{N-1}.
\]

Then we have the following inferences:

\[
0 \leq Pr\{E_1\} \leq Pr\{E_2\} \quad \text{by (29)}
\]

\[
= Pr\{\bigcup_{k=1}^{N-1} E_3^{(k)}\} \quad \text{by the definitions of} \ E_2 \ \text{and} \ E_3^{(k)}
\]

\[
\leq \sum_{k=1}^{N-1} Pr\{E_3^{(k)}\} = 0, \quad \text{by the union bound and (31)}
\]

i.e., \( Pr\{E_1\} = 0 \). Therefore, the proof is completed.
7. REFERENCES


