CONVEX GEOMETRY BASED OUTLIER-INSENSITIVE ESTIMATION OF NUMBER OF ENDMEMBERS IN HYPERSPECTRAL IMAGES

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

[†] Institute of Communications Engineering, National Tsing Hua Univ., Hsinchu, Taiwan
[*] Advanced Digital Sciences Center, Singapore
E-mail: `{aareul,thchan}@ieee.org`; `cychi@ee.nthu.edu.tw`

ABSTRACT

Accurate estimation of number of endmembers in a given hyperspectral data plays a vital role in effective unmixing and identification of the materials present over the scene of interest. The estimation of number of endmembers, however, is quite challenging due to the inevitable combined presence of noise and outliers. Recently, we have proposed a convex geometry based algorithm, namely geometry based estimation of number of endmembers - affine hull (GENE-AH) [1] to reliably estimate the number of endmembers in the presence of only noise. In this paper, we will demonstrate that the GENE-AH algorithm can be suitably used for reliable estimation of number of endmembers even for data corrupted by both outliers and noise, without any prior knowledge about the outliers present in the data. Initially, the GENE-AH algorithm (alongside with its inherent endmember extraction algorithm: p-norm-based pure pixel identification (TRI-P) algorithm) is used to identify the set of *candidate pixels* (possibly including the outlier pixels) that contribute to the affine dimension of the hyperspectral data. Inspired by the fact that the affine hull of the hyperspectral data remains intact for any data set associated with the same endmembers (that may not be in the data set), using GENE-AH again on the corrupted data with the identified candidate pixels removed, will yield a reliable estimate of the true affine dimension (number of endmembers) of that given data. Computer simulations under various scenarios are shown to demonstrate the efficacy of the proposed methodology.

***Index Terms*—** Estimation of number of endmembers, Hyperspectral data, Outliers, Affine hull, Convex geometry

1. INTRODUCTION

Hyperspectral Remote Sensing (HRS) is a powerful technology to record the electromagnetic reflections of an object, in hundreds of narrowly spaced spectral bands, so as to gather information about the object without getting in direct contact with it. However, the mixed pixel nature of the HRS image, demands effective hyperspectral unmixing (HU) to yield the details about the underlying objects [2]. Estimation of number of sources (aka model order selection) is the prime important processing step for HU, as it directly dictates the accuracy of the subsequent steps (dimension reduction, endmember extraction, and abundance estimation) and thereby, the accuracy of mineral identification and quantification. The inevitable presence of noise and outliers in the hyperspectral data poses a tough challenge for accurate estimation of number of endmembers. The noise in the hyperspectral data can result from the sensor noise and random atmospheric fluctuations and absorptions. On the other hand, the outliers can be simply regarded as pixels that deviate markedly from the rest of the pixels in the data, and such outliers may be either due to detector failure or data transfer errors or simply due to the presence of odd (unintended) endmembers in the given scene [3], [4].

Among the existing algorithms for estimation of number of endmembers, hyperspectral subspace identification (HySiMe) [5] and geometry based estimation of number of endmembers (GENE) [1], [6] have been shown to have promising performances. However, both HySiMe and GENE were proposed to estimate the number of endmembers in the presence of noise, and they may fail miserably when the data are also corrupted by outliers (refer to Table 2 in Section 4). Outlier mitigation is very challenging and our recently proposed algorithms to this end also requires the prior knowledge about the number of endmembers [7], [8]. So far, to the best of our knowledge, no existing algorithms can effectively estimate the number of endmembers from the hyperspectral data corrupted by both noise and outliers.

In this work, we show that the recently proposed geometry based estimation of number of endmembers - affine hull (GENE-AH) [1] can be suitably used to reliably estimate the number of endmembers from hyperspectral data simultaneously corrupted by both noise and outliers. It is found that the affine hull of the hyperspectral data, on which the GENE-AH algorithm is based, remains invariant for any data set associated with the same endmembers (that may not be in the data set). Because outliers can be treated as "pseudo endmembers", the GENE-AH is first used to detect the so-called *candidate pixels*, which includes true endmembers (pure pixels) and outliers, if any. Then all the detected candidate pixels are removed from the hyperspectral data and the remaining pixels are used to estimate the true number of endmembers, again by using the GENE-AH algorithm. Toy examples and some simulation results for various scenarios are provided to demonstrate the *feasibility* and *reliability* of the proposed methodology.

Notations: $\mathbb{R}^M$ and $\mathbb{R}^{M\times N}$ represent the set of real $M \times 1$ vectors and $M \times N$ matrices, respectively, $\mathbf{0}_M$ represents an $M \times 1$ all-zero vector, $\mathbf{1}_M$ represents an $M\times 1$ all-one vector, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. A Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted as $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. $\operatorname{aff}\{\mathbf{a}_1,\ldots,\mathbf{a}_N\}$ denotes the affine hull [9] of $\{\mathbf{a}_1,\ldots,\mathbf{a}_N\}$ and it is defined as

$$\operatorname{aff}\{\mathbf{a}_1,\ldots,\mathbf{a}_N\} = \left\{\mathbf{x} = \sum_{i=1}^{N}\theta_i\mathbf{a}_i \,\middle|\, \sum_{i=1}^{N}\theta_i = 1, \theta_i \in \mathbb{R},\ \forall i\right\}.$$

2. SIGNAL MODEL AND DIMENSION REDUCTION

The linear mixing model has been widely applied to model the HRS image data cube, by which each M-dimensional pixel vector (pixel

---

This work is supported by National Science Council (R.O.C.) under Grant NSC 99-2221-E-007-003-MY3.

for convenience) can be represented as [2], [10]:

$$\mathbf{x}[n]=\mathbf{A}\mathbf{s}[n]=\sum_{i=1}^{N}s_i[n]\mathbf{a}_i,\ \forall n=1,\ldots,L. \tag{1}$$

In (1), $\mathbf{x}[n]\in\mathbb{R}^M$ represents the nth pixel, $\mathbf{A}=[\ \mathbf{a}_1,\ldots,\mathbf{a}_N\ ]\in\mathbb{R}^{M\times N}$ denotes the endmember signature matrix with the ith column vector $\mathbf{a}_i$ being the ith endmember signature, $\mathbf{s}[n]\in\mathbb{R}^N$ is the nth abundance vector comprising N (usually in tens) fractional abundances, and L (usually in thousands) is the total number of observed pixels. However, noise and outliers are inevitable in reality, and hence the nth corrupted observed pixel $\mathbf{y}[n]\in\mathbb{R}^M$ can be expressed as:

$$\mathbf{y}[n]=\mathbf{x}[n]+\mathbf{w}[n]+\mathbf{z}[n],\ n=1,\ldots,L, \tag{2}$$

where $\mathbf{x}[n]$ is given by (1), $\mathbf{w}[n]\sim\mathcal{N}(\mathbf{0}_M,\sigma^2\mathbf{I}_M)$ in which σ^2 is the noise variance, and $\mathbf{z}[n]$ denotes the outlier vector which only appears at Z pixels, i.e.,

$$\begin{aligned}&\mathbf{z}[n]\neq\mathbf{0}_M,\ n\in\mathcal{I}\triangleq\{\ell_1,...,\ell_Z\},\\&\mathbf{z}[n]=\mathbf{0}_M,\ n\in\mathcal{L}\setminus\mathcal{I},\end{aligned}$$

where $\setminus$ is the set difference operator, $\mathcal{L}=\{1,2,...,L\}$, and $\mathcal{I}$ is the set of outlier pixel indices. Furthermore, it is assumed that the Z outliers are spectrally different from each other and each endmember contributes at least, to more than one pixel. It should be mentioned that the proposed methodology in this paper does not require any prior knowledge about the number of outliers (Z) nor its locations. Standard assumptions pertaining to the signal model in (1) are [2]: (A1) $s_i[n]\geq 0,\ \forall i,n$; (A2) $\sum_{i=1}^{N}s_i[n]=1,\forall n$; (A3) $M\geq N$, and $\mathbf{A}$ is of full column rank.

The dimension reduction procedure not only aids in considerable noise mitigation but also eases the computational load of the subsequent processes. Let $N_{\max}$ be an integer satisfying $N+Z\leq N_{\max}\leq M$. Such $N_{\max}$ is always possible, as in advanced HRS sensors the number of outliers are about 2 to 5 percent of L (which is usually very large) and M is about several hundreds. Following the discussions in [1], the corrupted dimension reduced observations can be obtained as:

$$\tilde{\mathbf{y}}[n]=\boldsymbol{\mathcal{C}}^T(\mathbf{y}[n]-\mathbf{d})\in\mathbb{R}^{N_{\max}-1}, \tag{3}$$

where

$$\mathbf{d}=\frac{1}{L}\sum_{n=1}^{L}\mathbf{y}[n], \tag{4}$$

$$\boldsymbol{\mathcal{C}}=[\ \boldsymbol{q}_1(\mathbf{U}_y\mathbf{U}_y^T),\ldots,\boldsymbol{q}_{N_{\max}-1}(\mathbf{U}_y\mathbf{U}_y^T)\ ], \tag{5}$$

in which $\mathbf{U}_y=[\ \mathbf{y}[1]-\mathbf{d},\ldots,\mathbf{y}[L]-\mathbf{d}\ ]$ is the mean removed data matrix and $\boldsymbol{q}_i(\mathbf{X})$ is the eigenvector associated with the ith largest eigenvalue of a matrix $\mathbf{X}$. Next, we present how GENE-AH can be suitably used to reliably estimate N from $\tilde{\mathbf{y}}[n],\ \forall n$.

3. ESTIMATION OF NUMBER OF ENDMEMBERS USING GENE-AH FOR NOISE-OUTLIER CORRUPTED DATA

The GENE-AH algorithm [1] is based on the data geometry fact that, regardless of the existence of pure pixels (pure pixel is a pixel contributed fully by a single endmember) in the given data set, all the noise-free observed pixel vectors should lie in the affine hull (AH) of the endmember signatures and its affine dimension is $N-1$. For the sake of self containment and clarity of presentation, we first briefly review the GENE-AH algorithm [1] based on the noise and outlier free signal model given by (1).

3.1. Brief Review of GENE-AH

The GENE-AH exploits the successive estimation property of a successive endmember extraction algorithm (EEA), namely p-norm-based pure pixel identification (TRI-P) algorithm. It aims to decide when the TRI-P algorithm should stop extracting the next new endmember. Let $\tilde{\mathbf{x}}[n]$ denote the noise-free dimension reduced pixel associated with $\tilde{\mathbf{y}}[n],\ \forall n$, and $l_1,\ldots,l_{k-1},l_k$ denote the purest pixel indices (i.e., candidate pixel indices) obtained by processing $\tilde{\mathbf{y}}[n],\ \forall n$ using the TRI-P algorithm. Given the dimension-reduced noisy hyperspectral data (defined by (2) with $\mathbf{z}[n]=\mathbf{0},\ \forall n$), to account for the noise in the hyperspectral data, Neyman-Pearson hypothesis testing [11] has been employed to find the smallest k, with a given false alarm probability P_{FA}. Precisely, at the kth recursion, the following constrained least squares (convex) problem is solved:

$$\boldsymbol{\theta}^\star=[\theta_1^\star,\ldots,\theta_{k-1}^\star]^T\triangleq\arg\min_{\mathbf{1}_{k-1}^T\boldsymbol{\theta}=1}\ \|\tilde{\mathbf{y}}[l_k]-\mathbf{A}_{k-1}\boldsymbol{\theta}\|_2^2,$$

where $\mathbf{A}_{k-1}=[\tilde{\mathbf{y}}[l_1],\ldots,\tilde{\mathbf{y}}[l_{k-1}]]$. Let

$$\boldsymbol{e}=\tilde{\mathbf{y}}[l_k]-\mathbf{A}_{k-1}\boldsymbol{\theta}^\star, \tag{6}$$

be the optimal fitting error, and it leads to the following binary hypotheses:

- H_0: $\tilde{\mathbf{x}}[l_k]\in\operatorname{aff}\{\tilde{\mathbf{x}}[l_1],\ldots,\tilde{\mathbf{x}}[l_{k-1}]\}$, and $\boldsymbol{e}\sim\mathcal{N}(\mathbf{0},\gamma)$.
- H_1: $\tilde{\mathbf{x}}[l_k]\notin\operatorname{aff}\{\tilde{\mathbf{x}}[l_1],\ldots,\tilde{\mathbf{x}}[l_{k-1}]\}$, and $\boldsymbol{e}\sim\mathcal{N}(\boldsymbol{\mu}_k,\gamma)$,

where $\boldsymbol{\mu}_k$ is unknown and $\gamma=(1+\theta_1^{\star 2}+\theta_2^{\star 2}+\cdots+\theta_{k-1}^{\star 2})\sigma^2$. GENE-AH employs a Neyman-Pearson hypothesis testing classifier [11] with a pre-defined P_{FA}, to obtain the estimate $\hat{N}=\hat{k}-1$ where $\hat{k}$ is the smallest integer such that hypothesis H_0 is decided by the Neyman-Pearson classifier. GENE-AH performs well in the presence of only noise [1], [6]. As the presence of outliers adds to the intrinsic affine dimension of the data, $\hat{N}$ turns out to be the sum of N and Z, due to that every outlier has been treated as an endmember by GENE-AH. Next, let us present how to estimate N in the presence of outliers (and noise) using GENE-AH.

3.2. Outlier-GENE-AH Algorithm

The idea behind the proposed methodology is to identify and remove all the candidate pixels for endmembers and outliers from the given data using GENE-AH, and then estimate the number of endmembers using the candidate pixels removed data set, again by using GENE-AH. This idea can be illustrated by the following toy example shown in Figures 1 - 3. Figure 1 shows the 2-dimensional scatter plot of the noise and outlier corrupted hyperspectral data where there are 3 endmembers ($N=3$) randomly chosen from USGS library [12], with no pure pixels in the data cloud, $M=224$ spectral bands, $L=1000$, and $Z=30$ (which amounts to 3% of L), SNR=40 dB, and signal-to-outlier ratio (SOR)=10 dB (to be defined in Section 4). The true endmembers are not present in the data cloud, but are marked for reference (red colored hollow triangles). When GENE-AH is applied to the above hyperspectral data with $N_{\max}=50$ and $P_{\mathrm{FA}}=10^{-6}$, the estimated number of endmembers is $\hat{N}=33$, which obviously is not an estimation of N but $N+Z$, and let the set of identified pixel indices be $\mathcal{I}_{\mathrm{est}}=\{l_1,l_2,\ldots,l_{\hat{N}}\}$. It is interesting (but not surprising) to note that the estimated number is the sum of the true number of endmembers and the number of outliers. This is because of the fact that the *outliers add to the original affine dimension of the hyperspectral data*. The candidate pixels estimated by the GENE-AH (along with its inherent EEA, TRI-P algorithm) are shown in Figure 2.

It is obvious to see from Figure 2 that the outliers pixels are now identified along with few other pixels (not necessarily outside

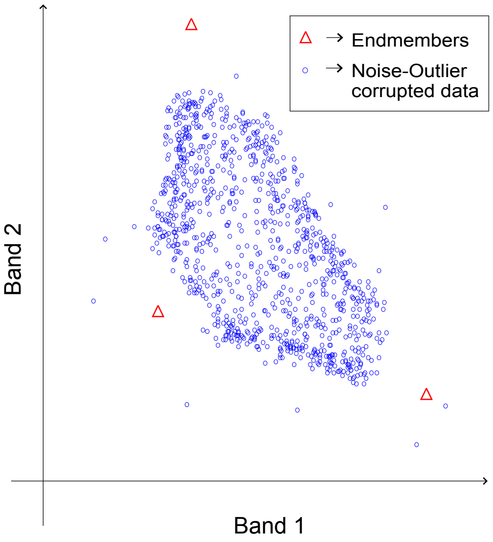

Fig. 1. Simulated noise-outlier corrupted hyperspectral data with $N = 3$, $L = 1000$, $Z = 30$, SNR=40 dB, and SOR=10 dB.

the data cloud). As the affine hull of the data remains intact even some pixels are removed, we then applied the GENE-AH (again with $N_{\max} = 50$) for the data set with the candidate pixels removed from the original data cloud. The data cloud with the candidate pixels removed is shown in Figure 3. The number of endmember is now estimated to be $\hat{N} = 3$, which is exactly the true number of endmembers.

In general, the data set after removing the candidate pixels will have $L - \hat{N}$ vectors, and can be expressed as:

$$\widetilde{\boldsymbol{\mathcal{Y}}} = \{\tilde{\mathbf{y}}[1], \ldots, \tilde{\mathbf{y}}[L]\} \setminus \{\tilde{\mathbf{y}}[l_1], \ldots, \tilde{\mathbf{y}}[l_{\hat{N}}]\}. \tag{7}$$

As the potential outliers are removed (only the noise being the source of data contamination), the effective and efficient GENE-AH (as extensively investigated in [1]), when applied to $\widetilde{\boldsymbol{\mathcal{Y}}}$, will yield accurate estimate of the number of endmembers. As the above procedure is based on the GENE-AH algorithm, it is named as Outlier-GENE-AH (O-GENE-AH) algorithm, which is summarized in Table 1.

Table 1. Pseudo-code of O-GENE-AH algorithm.

Given	noise and outlier corrupted hyperspectral data $\mathbf{y}[n]$, $\forall n$, $N_{\max}$ (between $N + Z$ and M), false alarm probability $P_{\rm FA}$ (e.g., 10^{-6}).
Step 1.	Compute $(\boldsymbol{\mathcal{C}}, \mathbf{d})$ given by (4) and (5), and obtain the dimension reduced observations $\tilde{\mathbf{y}}[n]$, $\forall n$ by (3).
Step 2.	Apply GENE-AH algorithm to $\tilde{\mathbf{y}}[n]$, $\forall n$ and let the estimated number of endmembers be $\hat{N}$, and let the set of identified candidate pixel indices be $\mathcal{I}_{\rm est} = \{l_1, l_2, \ldots, l_{\hat{N}}\}$.
Step 3.	Remove the candidate pixels corresponding to the pixel indices obtained in Step 2, to get $\widetilde{\boldsymbol{\mathcal{Y}}}$ (given by (7)).
Step 4.	Apply GENE-AH algorithm to $\widetilde{\boldsymbol{\mathcal{Y}}}$ and the estimated number of endmembers will be the desired estimate of the true number of endmembers.

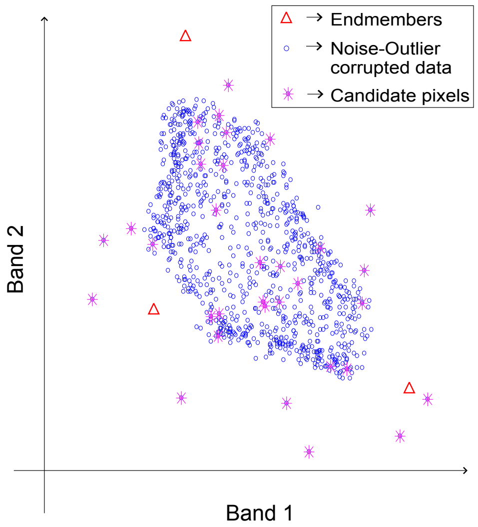

Fig. 2. The 33 candidate pixels identified by GENE-AH for the simulated noise-outlier corrupted hyperspectral data shown in Figure 1.

4. SIMULATION AND CONCLUSION

In this section the proposed O-GENE-AH for estimating the number of endmembers in the presence of both noise and outliers is evaluated using Monte Carlo simulations of 100 independent runs for various possible realistic scenarios. As demonstrated in [1] and [6], HySiMe [5] is one of the most promising existing algorithms, and hence it is used for performance comparison, besides the original GENE-AH algorithm. For the algorithms under test, for each simulated data set the true noise covariance matrix is used, and except for HySiMe the $P_{\rm FA} = 10^{-6}$.

The noise-free pixels are generated following the signal model in (1), where the endmember signatures with 224 spectral bands ($M = 224$) are chosen from the U.S. geological survey (USGS) library [12] and the abundance vectors are generated following the Dirichlet distribution which naturally accounts for (A1) and (A2). Another parameter of interest is the purity level ρ of the data set [1] which is defined as $\rho = \max\{\|\mathbf{s}[n]\|_2, n = 1, \ldots, L\}$. The smaller the value of ρ, the more mixed is the data set, and $\rho = 1$ implies the existence of pure pixels (i.e., true endmembers) in the data set. Randomly generated zero-mean white Gaussian noise vectors are added to the generated noise-free pixels, for various signal-to-noise ratios (SNRs) defined as

$$\text{SNR} = \frac{1}{\sigma^2 ML} \sum_{n=1}^{L} \|\mathbf{x}[n]\|^2. \tag{8}$$

The Z outlier pixel locations $\{\ell_1, ..., \ell_Z\}$ are randomly chosen from $\{1, \ldots, L\}$, and they are generated by [8]: $\mathbf{z}[\ell_i] = c\boldsymbol{\kappa}_i$, $i = 1, \ldots, Z$, in which each element of $\boldsymbol{\kappa}_i$ is a zero-mean unit-variance Laplacian random variable, and c is a scalar adjusted according to the predefined signal-to-outlier ratio (SOR), where

$$\text{SOR} = \frac{\sum_{n=1}^{L} \|\mathbf{x}[n]\|_2^2 / L}{\sum_{i=1}^{Z} \|\mathbf{z}[\ell_i]\|_2^2 / Z}. \tag{9}$$

As a general rule, in all the simulation scenarios SNR $\geq$ SOR, so as to ensure that the outlier in a given location has heavier contamination in that observed pixel than the noise. On the other hand

Table 2. Mean±standard deviation of the estimated number of endmembers for various scenarios over 100 independent runs.

Scenarios ($N = 8$, $L = 1000$)	HySiMe		GENE-AH		O-GENE-AH		O-GENE-AH2	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
$Z = 20$, SNR=25dB, SOR=10dB, $\rho = 0.8$	28.00	0.00	27.46	0.64	7.77	0.53	7.84	0.54
$Z = 20$, SNR=30dB, SOR=10dB, $\rho = 0.8$	28.00	0.00	28.00	0.00	8.02	0.14	8.02	0.14
$Z = 20$, SNR=35dB, SOR=10dB, $\rho = 0.8$	28.00	0.00	28.00	0.00	8.00	0.00	8.00	0.00
$Z = 20$, SNR=40dB, SOR=10dB, $\rho = 0.8$	28.00	0.00	28.00	0.00	8.00	0.00	8.00	0.00
$Z = 20$, SNR=30dB, SOR=15dB, $\rho = 0.8$	28.00	0.00	28.00	0.00	8.02	0.14	8.02	0.14
$Z = 20$, SNR=30dB, SOR=20dB, $\rho = 0.8$	27.97	0.17	28.00	0.00	8.03	0.17	8.02	0.14
$Z = 20$, SNR=30dB, SOR=10dB, $\rho = 0.9$	28.00	0.00	28.00	0.00	8.00	0.00	8.00	0.00
$Z = 20$, SNR=30dB, SOR=10dB, $\rho = 1.0$	28.00	0.00	28.00	0.00	8.00	0.00	8.02	0.20
$Z = 50$, SNR=30dB, SOR=15dB, $\rho = 0.8$	58.00	0.00	57.88	0.32	8.06	0.23	8.03	0.17
$Z = 70$, SNR=30dB, SOR=15dB, $\rho = 0.8$	78.00	0.00	77.50	0.59	8.09	0.28	8.13	0.33

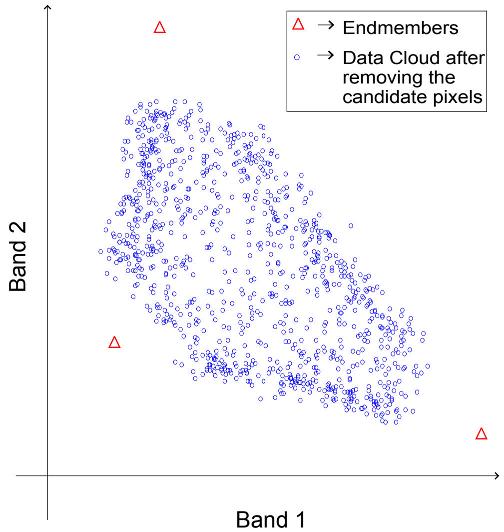

Fig. 3. Data cloud with the candidate pixels removed.

if SNR $<$ SOR, then the outliers just behave as noise vectors, and hence cannot be identified. The various scenarios that are considered for the purpose of evaluation are detailed in Table 2. The mean and the standard deviation of the estimated number of endmembers over 100 independent runs for each scenario is the performance measure. In Table 2, GENE-AH corresponds to applying GENE-AH [1] for the entire noise and outlier corrupted data, O-GENE-AH corresponds to applying GENE-AH to the data after removing the candidate pixels identified by GENE-AH. Just to ensure the reliability of O-GENE-AH, GENE-AH is, again, applied to the data after removing both the candidate pixels identified by GENE-AH and the candidate pixels identified by O-GENE-AH, and the results are tabulated under O-GENE-AH2 in Table 2. As can be readily inferred from Table 2, applying HySiMe and GENE-AH directly to the outlier-noise corrupted data resulted in the estimation of $N + Z$. After removing the candidate pixels from the original data cloud, the results of the estimated number of endmembers by the proposed O-GENE-AH are close to that of true number of endmembers (i.e., $N = 8$), and so are the results of the estimated number of endmembers (under the column O-GENE-AH2) by removing the candidate pixels twice from the original data set.

In conclusion, we have presented a methodology (O-GENE-AH, in Table 1) to reliably estimate the number of endmembers when the data are corrupted even by both noise and outliers. The recently proposed GENE-AH is used to identify and remove the so-called candidate pixels (consisting of purest pixels that include outliers) from the noise and outlier corrupted data cloud, and then it is applied again to that data cloud (with candidate pixels removed) to yield a high fidelity estimate of the number of endmembers. Simulation results demonstrated the efficacy of the proposed O-GENE-AH methodology. Testing O-GENE-AH on real AVIRIS hyperspectral data is currently under investigation. Considering other types of noise (non-Gaussian) and other types of outliers (non-Laplacian) will be our future direction.

5. REFERENCES

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

[2] 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 in Applied Earth Obs. and Remote Sensing*, vol. 5, no. 2, pp. 354-379, June 2012.

[3] 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.

[4] N. Aceto, M. Diani, and G. Corsini,“Hyperspectral signal subspace identification in the presence of rare signal components,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940-1954, April 2010.

[5] J. M. B. Dias and J. M. P. Nascimento, “Hyperspectral subspace identification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435-2445, Aug. 2008.

[6] A. Ambikapathi, T.-H. Chan, and C.-Y. Chi, “Convex geometry based estimation of number of endmembers in hyperspectral images,” in *Proc. IEEE ICASSP*, Kyoto, Japan, Mar. 25-30, 2012, pp. 1233-1236.

[7] 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,” to appear in *IEEE Trans. Geosci. Remote Sens.*

[8] H.-E. Huang, T.-H. Chan, A. Ambikapathi, W.-K. Ma, and C.-Y. Chi, “Outlier-robust dimension reduction and its impact on hyperspectral endmember extraction,” in *Proc. IEEE WHISPERS*, Shanghai, China, June 5-7, 2012.

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

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

[11] L. C. Ludeman. *Random Processes Filtering, Estimation, and Detection.* Wiley-Interscience Publication, 2003.

[12] R. N. Clark, G. A. Swayze, A. Gallagher, T.V. King, and W. M. Calvin, “The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0,” *U. S. Geol. Surv., Open File Rep.* pp. 93-592, 1993.