Determining the Intrinsic Dimension of a Hyperspectral Image Using Random Matrix Theory

Kerry Cawse-Nicholson, Steven B. Damelin, Amandine Robin, and Michael Sears, *Member, IEEE*

***Abstract*— Determining the intrinsic dimension of a hyperspectral image is an important step in the spectral unmixing process and under- or overestimation of this number may lead to incorrect unmixing in unsupervised methods. In this paper, we discuss a new method for determining the intrinsic dimension using recent advances in random matrix theory. This method is entirely unsupervised, free from any user-determined parameters and allows spectrally correlated noise in the data. Robustness tests are run on synthetic data, to determine how the results were affected by noise levels, noise variability, noise approximation, and spectral characteristics of the endmembers. Success rates are determined for many different synthetic images, and the method is tested on two pairs of real images, namely a Cuprite scene taken from Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) and SpecTIR sensors, and a Lunar Lakes scene taken from AVIRIS and Hyperion, with good results.**

***Index Terms*— Hyperspectral, intrinsic dimension, linear mixture model, random matrix theory, unmixing.**

I. Introduction

DETERMINING the number of sources in a signal is important for the processing of many different types of data, including chemical unmixing [1], extracting speech signals in a noisy line [2], unmixing minerals [3] and unmixing environmental landscapes [4], among many others. Determination of this number is necessary for classification methods, unmixing methods and target detection [5]. It is common practice for the user to select the number of endmembers to suit the application, but this does not necessarily agree with the intrinsic dimension of the image, and an incorrect estimation of this number may have detrimental effects on the end results [6].

Manuscript received July 20, 2011; revised October 24, 2012; accepted October 25, 2012. This work was supported in the part by the Centre for High Performance Computing flagship project: Computational Research Initiative in Imaging and Remote Sensing and CSIR. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian D. Rigling.

K. Cawse-Nicholson is with the School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2000, South Africa, and also with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria 0001, South Africa (e-mail: cawse@cis.rit.edu).

S. B. Damelin is with the Department of Mathematics, Wayne Country Day School, Goldsboro, NC 27530 USA (e-mail: steve.damelin@gmail.com).

A. Robin is with the School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2000, South Africa (e-mail: amandine.c.robin@gmail.com).

M. Sears is with the School of Computer Science, University of the Witwatersrand, Johannesburg 2000, South Africa (e-mail: michael.sears@wits.ac.za).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2227765

In this paper, we will focus on the application to hyperspectral imagery, although this does not preclude applicability to other areas. A hyperspectral image may be visualised as a 3-dimensional data cube of size ($m \times n \times p$). This is a set of p images of size $N = m \times n$ pixels, where all images correspond to the same spatial scene, but are acquired at p different wavelengths.

In a hyperspectral image, the number of bands, p, is large (typically $p \approx 200$). This high spectral resolution enables separation of substances with very similar spectral signatures. However, even with high spatial resolution, each pixel is often a mixture of pure components.

It is of interest to unmix each pixel, to determine the abundances of certain pure substances in the image. In unsupervised remote sensing, the spectra of the K pure substances themselves are also unknown, and the presence of noise forces the dataset to full dimension p. In practice the data set will involve significant redundancy. If we suppose that the data are noise free, then they would be contained in a proper vector subspace $\mathbb{R}^K$ of $\mathbb{R}^p$. Various terms have been introduced in the literature for similar concepts.

Bioucas-Dias and Nascimento [7] define the Intrinsic Dimension of the image (ID) as the dimension of the signal subspace. Chang and Du [8] define ID as the "minimum number of parameters required to account for the observed properties of the data." They also define a separate estimate, called *Virtual Dimension* (VD), which is the number of endmembers necessary to give accurate unmixing, which may be larger than the number of so-called "idealized substances." Here substances may be understood as pure targets or endmembers, and could represent different objects depending on the application, e.g. chemical powders, cover types such as tree species, etc. By definition, VD may be dependent on the unmixing method that is used, unlike ID. Wu *et al.* have compared estimates of VD and ID in their survey paper [6] and found them to be comparable for the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) scene of Cuprite.

Bajorski [9], however, found that the concept of VD may be misleading in certain circumstances, since the value may change when the image is shifted and rotated. Bajorski [9] instead defines the Effective Dimensionality (ED) as "the dimensionality of the affine subspace giving an acceptable approximation to all pixels." Schlamm *et al.* [10] define an Inherent Dimension K', where "the entire spectral image can lie in the same K'-dimensional hyperplane." These authors also claim that the Inherent Dimension is not equivalent to

VD, or the number of spectral signatures present in the image. In contrast, the Spanning Dimension is defined as the "minimum number of basis vectors required to span the space" of pixel observations [10]. Schlamm *et al.* define many other types of dimensions which may apply to hyperspectral imagery, to display the confusion around the term. These authors also define their own intrinsic dimension as "the smallest number of parameters needed to contain all of the variability in the data through a mapping function." This differs from Bioucas-Dias and Nascimento's definition.

The many different definitions of Intrinsic Dimension may be confusing, and although some of the above definitions are closely connected, some definitions may lead to different results, and so it is core for this paper to state a formal definition for ID that is independent of the method used to calculate it.

First, we define some generic notation that will be used in this paper. Underlined variables, such as $\underline{x}$ represent vectors, random variables are denoted by a tilde, $\tilde{x}$, and subscript indices $\underline{x}_i$ refer to measurement $\underline{x}$ for the i^{th} sample.

For each pixel i, $1 \leq i \leq N$, let $\underline{x}_i \in \mathbb{R}^p$ be the observed spectral measurement. Assuming the measurement may be decomposed into signal and noise, write $\underline{x}_i = \underline{s}_i + \underline{\xi}_i$, where $\underline{s}_i$ represents the information in pixel i and $\underline{\xi}_i$ represents the noise. We introduce the following definition.

Definition 1 (Intrinsic Dimension): The Intrinsic Dimension (ID) of a dataset $[\underline{x}_1, \ldots, \underline{x}_N]$ is the dimension, K, of the vector subspace spanned by $\{\underline{s}_i, 1 \leq i \leq N\}$.

Definition 1 is considered to be equivalent to the Bioucas-Dias and Nascimento definition [7], since $\{\underline{s}_1, \ldots, \underline{s}_N\}$ spans the signal subspace, and therefore K is the dimension of the signal subspace. This is also equivalent to the spanning dimension defined in [10].

A. Literature Review

Some of the most popular methods to calculate the ID of a hyperspectral image include Maximum Orthogonal-Complements Algorithm (MOCA) [11], Harsanyi–Farrand–Chang (HFC) [8] and HySime [12]. HySime is a modification of Signal Subspace Estimation (SSE) [7]. MOCA is specifically designed to preserve rare substances and has been improved in terms of computational complexity by Acito *et al.* in [13]. This method has also been combined with HFC by Acito *et al.* [14] and Chang *et al.* [15]. MOCA assumes independent identically distributed (i.i.d.), Gaussian noise, but may be adapted for non-i.i.d. and correlated noise [11].

In 2007, Wu *et al.* [6] provided a summary of methods to determine the intrinsic dimension of a hyperspectral image. This paper compares the HFC Detection Method [16], which assumes white noise with zero mean, with a number of other methods which require various noise assumptions. All the methods discussed only use the eigenvalues of the observation covariance matrix. In all cases, nothing needs to be known about the basis vectors, which is an advantage over supervised methods. These methods are all taken from different areas, including chemistry and signal array processing, and are applied to hyperspectral imagery.

From the real and synthetic experiments described in [6], the authors determined that the best method for hyperspectral imagery was HFC, and the methods with the strictest noise assumptions (i.e. that the noise is Gaussian and i.i.d.) performed the worst. On the other hand, HFC was shown to be sensitive to user-defined values.

B. Linear Mixture Model

A common model used in remote sensing to separate mixed signals is the linear mixture model, introduced by Horwitz *et al.* [17]. This model assumes that the measurement in each pixel is made up of a convex linear combination of pure components. Kritchman and Nadler [1] define the set of pixel observations by the $(p \times N)$ matrix $X = [\underline{x}_1, \ldots, \underline{x}_N]$ and assume that for each pixel indexed by i, the observed measurement vector $\underline{x}_i \in \mathbb{R}^p$ is a realisation of a random vector $\underline{\tilde{x}}$, so that

$$\underline{\tilde{x}} = V\underline{\tilde{u}} + \underline{\tilde{\xi}}, \tag{1}$$

where $\underline{\tilde{u}}$ is a $(K \times 1)$ random vector subject to the constraints $\underline{\tilde{u}} \geq \underline{0}$ and $\underline{1}^T\underline{\tilde{u}} = 1$, and represents the proportions of pure components; V is a $(p \times K)$ deterministic matrix, with columns corresponding to the spectral measurements characterizing the pure components present in the whole scene; $\underline{\tilde{\xi}}$ is a random vector in $\mathbb{R}^p$ representing the noise; and K is the total number of pure components.

It is commonly assumed that $\underline{\tilde{\xi}}$ follows a Gaussian distribution $\mathcal{N}(0, \Phi)$, where Φ is the noise correlation matrix. This model allows the noise to be spectrally but not spatially correlated. This assumption has been successfully used in chemical unmixing [1] and in many methods described in the survey paper by Wu *et al.* [6]. The matrix Φ must be estimated from the data using a noise approximation method such as Meer's method [18].

C. Background on Random Matrix Theory

Determining the ID of a dataset is most often done by analyzing the eigenvalues of the observation covariance matrix, $S(N)$, defined as

$$S(N) := \frac{1}{N}\sum_{i=1}^{N} \underline{x}_i \underline{x}_i^T, \tag{2}$$

where N is the number of samples. Without loss of generality, centered data is assumed, so that the mean pixel value over the entire image is $\underline{0}$. The matrix $S(N)$ is used to distinguish the eigenvalues due to signal and the eigenvalues due to noise. This procedure is used to unmix hyperspectral images [6], signals [2], chemical mixtures [1] and others [3], [4].

In [1], $\underline{\tilde{x}}$ and $\underline{\tilde{\xi}}$ are independent, and $\Phi = \sigma^2 I_p$. This is reasonable in the chemical application being considered, since repeated samples are taken from the same mixture. The model covariance S is defined by $S = \mathbb{E}[\underline{\tilde{x}}\underline{\tilde{x}}^T]$. S may

be diagonalised by W so that

$$W^T S W = W^T V \mathbb{E}\left[\underline{\tilde{u}\tilde{u}}^T\right] V^T W + \sigma^2 I_p \tag{3}$$

$$= \begin{pmatrix} l_1 & & 0 \\ & \cdots & \\ & & l_K \\ 0 & & 0 \end{pmatrix} + \sigma^2 I_p \tag{4}$$

where l_i represents the i^{th} signal eigenvalue, and the observed eigenvalues tend toward $l_i + \sigma^2$ for $1 \leq i \leq K$ and to σ^2 for the remaining (noise) eigenvalues. The authors in [1] state that $S(N) \to S$ as $N \to \infty$ with probability one. Since in a hyperspectral image N is large, the eigenvalues of the sample covariance might be expected to show this clear distinction between eigenvalues due to signal and eigenvalues due to noise, allowing easy calculation of K. However, in practice, the values of the observed noise eigenvalues can be highly variable, making it difficult to distinguish between a small signal eigenvalue and a large noise eigenvalue. Several methods have been developed to address this problem.

In the method described in [1], the following assumptions are made: the columns of V are linearly independent; the abundance correlation matrix $\mathbb{E}[\underline{\tilde{u}\tilde{u}}^T]$ has full rank; and the noise is uncorrelated with the signal. In order to better understand the assumption that $\mathbb{E}[\underline{\tilde{u}\tilde{u}}^T]$ has full rank, consider an example where two vectors in V always occur in the same proportions, resulting in reduced rank of $\mathbb{E}[\underline{\tilde{u}\tilde{u}}^T]$. In this case, however, we would expect that the value of K for the scene would be lower than the number of pure substances. Thus this assumption that $\mathbb{E}[\underline{\tilde{u}\tilde{u}}^T]$ has rank K is reasonable.

In [1], Kritchman and Nadler have worked with new results in Random Matrix Theory (RMT) to determine which eigenvalues are due to noise and which are due to signal. Because Gaussian noise is assumed, the noise eigenvalues perform like eigenvalues of a random matrix. The distribution of the largest such eigenvalue (the first non-signal eigenvalue) has been well studied [19], [20], and this allows accurate determination of K. The advantage to this method is that there are no parameters that need to be set by the user. It achieved good results in chemical testing.

Wishart matrices play an important part in RMT. Johnstone [19] describes a random cross-product matrix $\tilde{A} = \tilde{X}\tilde{X}^T$, where $\tilde{X}$ contains N independent column vectors, each following a p-variate Gaussian distribution $\mathcal{N}(\mu, \Phi)$. Then $\tilde{A}$ has a p-variate Wishart distribution with N degrees of freedom, $W_p(\Phi, N)$. Johnstone [19] derives results for the case where the mean of each column is zero and the standard deviation of the p-variate Gaussian distribution is the identity matrix, *i.e.* each column of X follows the normal distribution $\mathcal{N}(0, I_p)$, and then $\tilde{A} \sim W_p(I_p, N)$.

Random matrices were first used in physics, to determine quantum energy levels [19]. In this setting, both p and N were large, breaking away from traditional statistics, where p was fixed. And so the authors in [1] assume that $N \to \infty$, $p \to \infty$, with $\frac{p}{N} \to c$, where $c > 0$ is constant. A large body of research in RMT has been dedicated to the distribution of the largest eigenvalue of a matrix following a Wishart distribution [1], [19].

Kritchman and Nadler [1] used some of Johnstone's results [19] in their chemical application to determine the largest noise eigenvalue, by determining the largest sample covariance eigenvalue consistent with the distribution of the largest eigenvalue from a Wishart matrix. This method assumes i.i.d. noise. According to Johnstone [19], the largest eigenvalue of such a real-valued Wishart matrix, $\tilde{\lambda}_1$, fulfills the following condition with convergence in distribution in the limit as $N, p \to \infty$, $p/N \to c > 0$:

$$Pr\{\tilde{\lambda}_1 \leq \sigma^2(R_\mu(N, p) + s(\alpha)R_\sigma(N, p))\} \sim TW, \tag{5}$$

where TW denotes the Tracy-Widom distribution, and the rate of convergence is $O(p^{-2/3})$ [1]. In (5), σ^2 is the variance of the Gaussian noise, α is a significance level and $s(\alpha)$ may be found by inverting the Tracy-Widom distribution (in [1], $\alpha = 0.5\%$, and after experimentally investigating the sensitivity to α, we determine that this is optimal and fix this value for all images investigated). The Tracy-Widom distribution is the solution of a second order Painlevé ordinary differential equation [1].

For real valued data define

$$R_\mu(N, p) = \frac{1}{N}\left(\sqrt{N - \frac{1}{2}} + \sqrt{p - \frac{1}{2}}\right)^2 \tag{6}$$

$$R_\sigma(N, p) = \frac{1}{N}\left(\sqrt{N - \frac{1}{2}} + \sqrt{p - \frac{1}{2}}\right) \times \left(\frac{1}{\sqrt{N - \frac{1}{2}}} + \frac{1}{\sqrt{p - \frac{1}{2}}}\right)^{1/3}. \tag{7}$$

Note that these functions do not depend on the intrinsic dimension, K. It is also important that $R_\sigma(N, p) \to 0$ as p, $N \to \infty$, $p/N = c$ fixed, so that the $s(\alpha)R_\sigma(N, p)$ term in (5) tends to zero as the image becomes larger. This means that, especially for large images, the formula is not sensitive to the choice of α.

Traditionally, statistical techniques consider the scenario p fixed, with $p << N$, and in this case, the observed noise eigenvalues should be very close to σ^2 in the case of i.i.d. noise [21]. However, although the number of pixels in a hyperspectral image are several orders of magnitude larger than the number of spectral bands, the dividing line between signal and noise eigenvalues is in practice still unclear. El Karoui [21] found that high dimensional problems (a hyperspectral image may contain hundreds of spectral bands) might be better solved by assuming that p and N are both large, with their ratio fixed. Similarly, Kritchman and Nadler [1] state that (5) still holds in the case of finite but large N and p, and in fact their algorithm deals specifically with small N, with good results.

There is, however, a limit to the size of the eigenvalue that may be successfully detected. The phase transition phenomenon described in [20] results in the limit λ_{crit}, below which the noise eigenvalue will not be successfully identified. This limit is defined as

$$\lambda_{crit} = \sigma^2\sqrt{\frac{p}{N}}. \tag{8}$$

The use of RMT in the chemical application proved to be very powerful in the circumstances investigated [1], but the experiments were limited to small numbers of samples, multiple measurements of the same mixture, and i.i.d. noise. This work has been continued by the same authors in [22], where Kritchman and Nadler showed that the method is applicable to a large number of samples in a signal processing environment.

However, the noise was still assumed to be i.i.d. across bands and across samples. Nadler and Johnstone [23] further pursued the idea by considering i.i.d. noise, but possibly correlated samples. They assume the existence of noise only observations and show that the data may be whitened and a variation of Roy's test may be applied.

Since some of the assumptions are not applicable to hyperspectral imagery, in this paper, we extend the RMT ideas from [1] so that they are appropriate for a wider range of applications. We consider large numbers of samples, where each sample may differ greatly from the sample mean, and Section II shows that RMT can be used to develop a new procedure that is applicable to hyperspectral images even in the presence of non-i.i.d. and spectrally correlated noise.

Section III shows the new method in algorithmic format, Section IV shows that this method performs well on synthetic data, and Section V shows that RMT is now applicable to real hyperspectral images. The advantage of the method presented is that it is not dependent on a user-determined parameter, as are other hyperspectral methods discussed above.

II. Model Formulation

Some of the most accurate methods described above have drawbacks in terms of noise assumptions or sensitivity to user-defined parameters. We introduce a method which only assumes Gaussian noise that is independent from the signal (the noise may be correlated across bands, however), and does not rely on user-defined parameters. Specifically, $V\underline{\tilde{u}}$ is independent from $\underline{\tilde{\xi}}$ in (1), and we replace the i.i.d. noise variance $\sigma^2 I_p$ used in [1] with a variance Φ that allows for correlated, non-i.i.d. noise. The core of the method is based on Random Matrix Theory (RMT), and in this section the model will be developed.

First we consider the threshold λ_{crit} in (8), below which noise eigenvalues will not be detected. In hyperspectral images, N is typically much larger than p, and we assume that the eigenvalues we test are above this limit. (Experimentally, in our images this has proven to be true.)

To investigate whether the method described in [1] is applicable to hyperspectral images, we consider a hyperspectral image of pure noise, *i.e.* the measurement $\underline{\tilde{x}}_i$ in each pixel follows a Gaussian distribution $\mathcal{N}(0, \Phi)$. We define $\tilde{X} = [\underline{\tilde{x}}_1, \ldots, \underline{\tilde{x}}_N]$, then as stated above, $\tilde{A} = \tilde{X}\tilde{X}^T$ follows a Wishart distribution $W_p(\Phi, N)$. Denoting $\tilde{S}(N) = N^{-1}\sum_{i=1}^{N} \underline{\tilde{x}}_i \underline{\tilde{x}}_i^T$, which may also be written as $\tilde{S}(N) = N^{-1}\tilde{X}\tilde{X}^T$, then $N\tilde{S}(N)$ follows a Wishart distribution $W_p(\Phi, N)$. Hence, $S(N)$ defined in (2) may be seen as a realisation of $\tilde{S}(N)$ in the case of a pure noise image.

The i.i.d. noise assumption in [1] is not applicable to hyperspectral images, and whitening methods were determined to be unsuccessful in a hyperspectral application [24]. Therefore a new method will be developed to allow for correlated and non-i.i.d. noise.

In [1], the separation between noise and signal eigenvalues is possible because the eigenvalues of S, where $S(N) \to S$ for $N \to \infty$, are given by

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \cdots & \\ & & \lambda_K \\ 0 & & 0 \end{pmatrix} + \sigma^2 I_p.$$

However, this is only true when the standard deviation of the noise is constant across bands. The method using constant noise standard deviation was applied directly to hyperspectral imagery in [25] and, although the method showed good results when applied to synthetic images, on a real image no eigenvalue was found to satisfy (5). This result was to be expected, since it is known that hyperspectral images do not have the same noise standard deviation in each band. Whitening methods were examined in [24], which decorrelate the noise and force the noise standard deviation to become constant in each band so that the original method should be applicable. This improved the previous results, but K was still overestimated. In this paper we will further develop the method so that it is applicable for non-i.i.d. noise, and in order to do that we must first prove that the eigenvalues are still separable in the non-i.i.d. case. To show this we derive Propositions 1 and 2 below.

Proposition 1: Suppose $\underline{\tilde{x}} \in \mathbb{R}^p$ is a random column vector described by $\underline{\tilde{x}} = V\underline{\tilde{u}} + \underline{\tilde{\xi}}$, where V is a $(p \times K)$ matrix with linearly independent columns, $\underline{\tilde{u}}$ is a random vector in $\mathbb{R}^K$, $\underline{\tilde{\xi}}$ is a random vector in $\mathbb{R}^p$, where $\underline{\tilde{\xi}} \sim \mathcal{N}(0, \Phi)$, and Φ is the noise correlation matrix.

Assuming that $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ has full rank K, and that $(V\underline{\tilde{u}})$ and $\underline{\tilde{\xi}}$ are independent, the expectation $S = \mathbb{E}[\underline{\tilde{x}}\underline{\tilde{x}}^T]$ may be decomposed as $S = \Pi + \Phi$, where Π is a symmetric matrix of rank K.

Proof: See Appendix VI-A. ■

Proposition 1 shows that under an assumption of independence, the signal and noise are separable in the correlation matrix, even if the noise correlation matrix Φ contains varying diagonal terms and/or off-diagonal terms i.e. the noise is not i.i.d.. Random Matrix Theory formulae rely on Φ having the form $\sigma^2 I_p$, so let us write Φ as $\sigma^2 I_p + \bar{\epsilon}$, where σ^2 is the average of the diagonal terms of Φ. Note that S and Φ are symmetric, resulting in symmetric Π. The following proposition proves that the eigenvalues of the observation covariance matrix are also separable, so that the RMT evaluation may be applied.

Proposition 2: Suppose S is a $(p \times p)$ symmetric matrix described by $S = \Pi + \Phi$, where Π and Φ are also symmetric. Suppose $\Phi = \sigma^2 I_p + \bar{\epsilon}$ (σ^2 is scalar). Let λ_i^S be the i^{th} eigenvalue of S (sorted in descending order) and λ_i^Π be the i^{th} eigenvalue of Π (sorted in descending order). Then, assuming

$(\underline{x}_i^{\Pi})^T \underline{x}_i^S \neq 0$

$$\lambda_i^S = \lambda_i^{\Pi} + \frac{(\underline{x}_i^{\Pi})^T \Phi \underline{x}_i^S}{(\underline{x}_i^{\Pi})^T \underline{x}_i^S} \tag{9}$$

$$= \lambda_i^{\Pi} + \sigma^2 + \frac{(\underline{x}_i^{\Pi})^T \bar{\epsilon} \underline{x}_i^S}{(\underline{x}_i^{\Pi})^T \underline{x}_i^S}, \tag{10}$$

where $\underline{x}_i^S$ and $\underline{x}_i^{\Pi}$ are the eigenvectors of S and Π, corresponding to λ_i^S and λ_i^{Π} respectively.

Proof: See Appendix VI-B. ∎

This result is analogous to the one obtained in the case where $\Phi = \sigma^2 I_p$ shown in (4), but with extra terms involving the eigenvectors of Π and S. Propositions 1 and 2 provide a new formula for the noise in the RMT evaluation. For all $1 \leq i \leq p$

$$\rho_i = \frac{(\underline{x}_i^{\Pi})^T \Phi \underline{x}_i^S}{(\underline{x}_i^{\Pi})^T \underline{x}_i^S}, \tag{11}$$

where ρ_i is the difference between the i^{th} observed eigenvalue of S and the i^{th} signal eigenvalue as described by Proposition 2. The original RMT theory in [19] is used to describe the distribution of the largest eigenvalue of a Random Matrix. The authors in [1] have shown that this theory may also be used to detect the largest noise eigenvalue, where the study assumed Gaussian and i.i.d. noise. Since we evaluate each eigenvalue individually, each may be thought of as the largest eigenvalue of a submatrix, where the submatrix of size $(q \times q)$, $q = 1, \ldots, p$, has eigenvalues corresponding to the q smallest eigenvalues of S. So the threshold condition for noise eigenvalues (5) may now be written, for $i = 1, \ldots, p$, as:

$$\lambda_i < \rho_i R,$$
$$\text{where} \quad R = (R_\mu(N, p) + s(\alpha) R_\sigma(N, p)). \tag{12}$$

Now ρ_i depends on knowledge of the noise correlation matrix Φ, the eigenvalues of Π (remember that the observation covariance matrix $S(N)$ approximates $\Pi + \Phi$), and the eigenvalues of $S(N)$. Note that every eigenvalue must be tested, since the value for ρ_i will differ for each evaluation.

In this section the theory has been derived to determine the ID of a hyperspectral image, using RMT. This method allows for non-i.i.d. and correlated noise and requires only an estimate of the noise covariance matrix.

The next section will show how this theory can be used in practice for the estimation of intrinsic dimension, giving a summary of the algorithm used.

III. Algorithm

In section II, we showed how eigenvalue evaluation could be used theoretically to discriminate noise from signal eigenvalues. In order to apply these theoretical results in practice, we need to compute the observation covariance matrix $S(N)$ from the data and approximate the noise correlation matrix Φ.

Note that the RMT significance level, α, may in principle be used as a free parameter as in [22], but we choose to fix this value for all images so that there is no user-determined parameter. This is reasonable because the RMT evaluation is not sensitive to α when images are large, since $s(\alpha)$ is multiplied by $R_\sigma(N, p)$, which tends to zero as $N \to \infty$ (see (5) and (7)). In [1], Kritchman and Nadler fixed alpha to be 0.5%. Using this value in experiments with pure noise images, containing low numbers of pixels (∼1,000) we found that 99.7% of the time all eigenvalues were ascribed to noise. On the other hand, if a single signal is included in the image, one eigenvalue was always ascribed to signal. From these experiments this is indeed a reasonable value for α for hyperspectral data and everywhere in this paper it will be assumed that $\alpha = 0.5\%$. It should be noted that α is not dependent on the image (as it is fixed for all images) and is not considered a tuneable parameter.

Algorithm 1

$S(N) \leftarrow$ observation covariance matrix
$\underline{\lambda} \leftarrow [\lambda_1, \ldots, \lambda_p]$, the sorted (descending) eigenvalues of $S(N)$

$p \leftarrow \text{length}(\underline{\lambda})$
$R_\mu \leftarrow \frac{1}{N}(\sqrt{N - 0.5} + \sqrt{p - 0.5})^2$
$R_\sigma \leftarrow \frac{1}{N}(\sqrt{N - 0.5} + \sqrt{p - 0.5})(\frac{1}{\sqrt{N-0.5}} + \frac{1}{\sqrt{p-0.5}})^{\frac{1}{3}}$
$s \leftarrow (-\frac{3}{2}\log[4\sqrt{\pi}\frac{\alpha}{100}])^{\frac{2}{3}}$
$R \leftarrow R_\mu + s \times R_\sigma$

$f \leftarrow$ call NoiseApproximationFunction(Image)
$\Phi \leftarrow$ diagonal matrix containing variances per band determined by f
$\sigma^2 \leftarrow$ mean of the noise variances over all bands *i.e.* mean($[\Phi_{1,1}, \ldots, \Phi_{p,p}]$)
$E_1 \leftarrow$ eigenvectors of $S(N)$ (sorted so that corresponding eigenvalues are descending)
$E_2 \leftarrow$ eigenvectors of $(S(N) - \Phi)$ (sorted so that corresponding eigenvalues are descending)

for $i = 1 \rightarrow p$ **do**
  $\rho_i \leftarrow \frac{E_1^i . \Phi . E_2^i}{E_1^i . E_2^i}$
end for
$\underline{\rho} \leftarrow [\rho_1, \cdots, \rho_p]$
$K \leftarrow$ intersection between $\underline{\lambda}$ and $\underline{\rho} \times R$

To approximate the noise, we chose an algorithm described by Meer *et al.* in [18]. This is a pyramid algorithm that searches for homogeneous areas, and it has various rules to allow for outliers. In test images, this algorithm was able to approximate the noise accurately with RMSE of 1%. This particular algorithm produces only the variance per band, so in this case Φ is diagonal. Since we are assuming this noise approximation method, from this point on we assume a diagonal noise covariance matrix. Note that this is not a restriction to our method, any preferred noise approximation algorithm may be used. Meer's method was chosen since it estimates the noise on a band-by-band basis, and thus it is independent of any relationship between bands, such as noise correlation. The sensitivity of our method with respect to the accuracy of the noise approximation will be tested and discussed in Section IV-B.

Unlike previous versions [1], [25], our method allows for non-i.i.d. noise. This procedure is summarised in algorithmic format above.

The algorithm requires only the image itself as an input, and returns the intrinsic dimension of the image, K. It requires no prior knowledge about the image, and requires no user-determined inputs. In the next section, the robustness of this approach is analysed with regards to each variable of the model, using simulated data.

IV. Performance Analysis

A. Description of the Test Dataset

We simulated a dataset using 20 minerals taken from the JPL spectral library [26]. At each iteration of the algorithm, K unique minerals are chosen randomly from the dataset, and the proportions of each of the basis vectors in each pixel are randomly selected with the only restrictions being the positive and sum-to-one conditions that are enforced. We also varied our values for p (number of bands), N (number of pixels), and σ (the average of noise standard deviations across all bands) in these images. As seen in Figure 1, the spectra in our dataset vary widely—some are similar and others easily separable. By mixing random spectra together we simulate images that may be easy or difficult to process. Since the proportions are randomly selected as well, one or more basis vector may occur in very small proportions in the image. All of these properties were created in order to simulate as realistic a dataset as possible.

While pure endmembers may be well understood in mineral applications, this idea becomes more subjective in data that contains vegetation spectra, which may vary according to species, phenology, etc. In this paper it is assumed that each target may be represented by a single spectral signature, where variation around this target is understood to be smaller in magnitude than the noise present in the image. If the variation is larger than this, then certain vegetation classes may be represented by more than one spectral signature.

In each test, we fixed all variables (except the one being examined) at the following default values: $N = 10,000$, $p = 200$, $\sigma = 10^{-3}$, $K = 5$.

When noise statistics are assumed known, the method is independent of N and p. In this case, the method also produces 100% success rates for non-i.i.d. noise and noise that is correlated between bands. Note that in all tests below, the actual noise values used to construct the synthetic dataset are known, so that we test the robustness of the ID method rather than the accuracy of any noise approximation method. If the noise is approximated, accuracy will decrease.

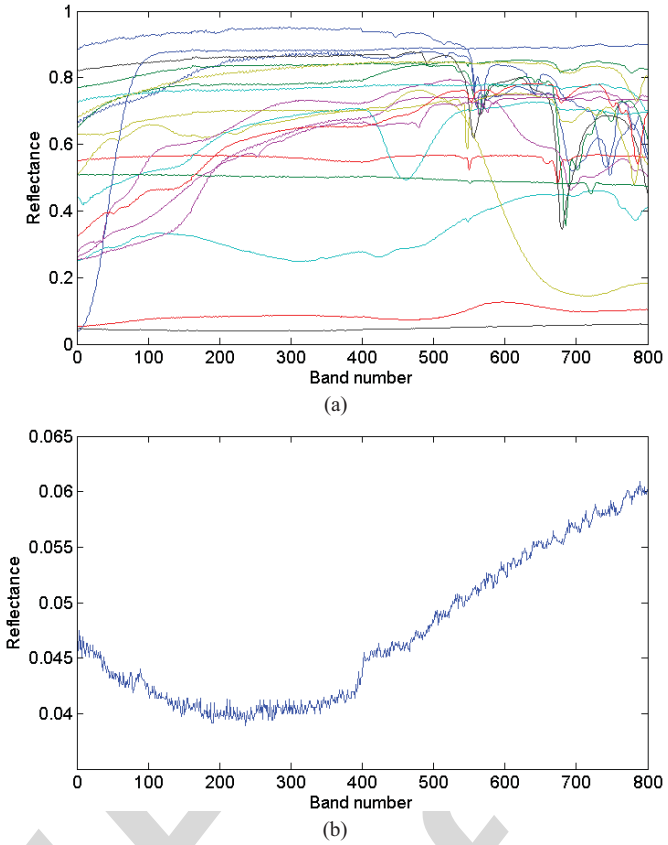


Fig. 1. (a) All the spectra used to make up the test data set. Note that there is a mix between similar and easily separable spectra, flat spectra and those with sharp features, spectra with high and low amplitude, etc. This is done to mimic a real environment. (b) Spectrum (Graphite 1A) with the lowest amplitude is shown at a larger scale. Take note of the noisiness of the spectrum.

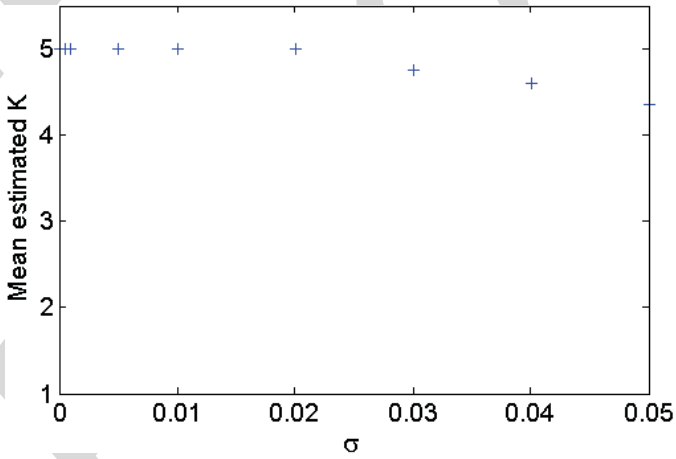


Fig. 2. Mean-estimated ID with regards to different noise levels. Twenty different images per noise level were simulated to calculate these values, and the true value for K is 5. Note that the method produces the correct value $K = 5$ in all 20 simulations for noise standard deviation up to 0.02.

B. Robustness to the Noise Level

Noise is an important element to test, as high noise eigenvalues can become very close to small signal values, making them difficult to separate. With low SNR (signal to noise ratio), two similar spectra will become indistinguishable. As seen in Figure 1, our test dataset does contain some basis vectors which are very close together, and so we are interested to test the level at which they are impossible to separate. Our tests in Figure 2 show that the mean estimated K, over 20 simulations, is exactly 5 for noise standard deviation below $\sigma = 0.02$. We use the SNR definition in [6] which says that the SNR is equal to half the mean signal divided by the standard deviation of the noise. Since our synthetic image mean is close to 0.5, a noise value of $\sigma = 0.01$ is equivalent to an SNR of 25:1, which is the limit of the noise tested in [8]. This lower bound is much lower than current hyperspectral sensors, with AVIRIS having an SNR of approximately 500:1 [3]. It should

be noted, however, that the intrinsic dimension will not be accurately determined for images containing very high levels of noise.

Note in Figure 1(a) that there are two spectra of a much lower amplitude than the rest. We will discuss these in Section IV-C, and study the impact of the number of basis vectors on the performance.

C. Performance With Regards to the Spectral Characteristics of Basis Vectors

Certain characteristics of basis vectors will make them difficult to detect in any model. Multiple spectra that have similar spectral signatures may be difficult to separate, and spectra with low amplitudes may be confused with noise. We have investigated the impact of such scenarios in our simulated dataset. Figure 1 clearly shows that our dataset contains spectra with different characteristics in order to mimic a real environment. Some spectra may appear to be similar, but our method was able to separate these. Two spectra stand out with very low amplitudes, and one spectrum in particular, Graphite, (Figure 1(b)) has low amplitude and high noise (mean reflectance 0.0475 and standard deviation 0.007). This spectrum was impossible to distinguish from noise with the default noise standard deviation of $\sigma = 10^{-3}$. This is equivalent to an SNR of 250:1, which is still low compared to real images (recall that AVIRIS has SNR = 500:1). Note that if the noise is reduced to $\sigma = 10^{-4}$, the basis vectors are separated with 100% accuracy. It is interesting that the other low-amplitude spectrum is still accurately detected—it may be the combination of the noisy signal and the low amplitude in Graphite that causes it to be recognised as noise in a noisy image.

D. Robustness to Noise Approximation

As previously stated, our algorithm does not depend on a specific noise approximation algorithm. However, the algorithm that is used must be accurate enough for our purposes. In this section, we test the impact of errors in noise approximation on the accuracy of our method in order to evaluate how accurate the noise approximation needs to be.

If we assume that the noise distribution in all bands is $\mathcal{N}(0, \sigma^2 I_p)$, Figure 3 shows that our algorithm only tolerates up to 10% underestimation, and overestimations of constant noise are well tolerated. This asymmetrical behaviour is firstly due to the eigenvalue spread (noise eigenvalues are closer together, so a noise approximation error towards the noise direction will result in higher errors in estimating K), and secondly due to the fact that the RMT evaluation in (5) is a single sided inequality; i.e. if the upper limit for the largest noise eigenvalue is increased, the noise eigenvalue will still fall below this limit. Conversely, if it is decreased, some noise eigenvalues will be recognised as signal. Again, this is made more predominant due to the asymmetrical spread of the eigenvalues. Therefore, for the best results, the noise approximation algorithm should be chosen with care, since the ID may not be accurate if noise is underestimated. Note that when the noise is spectrally correlated, it is often

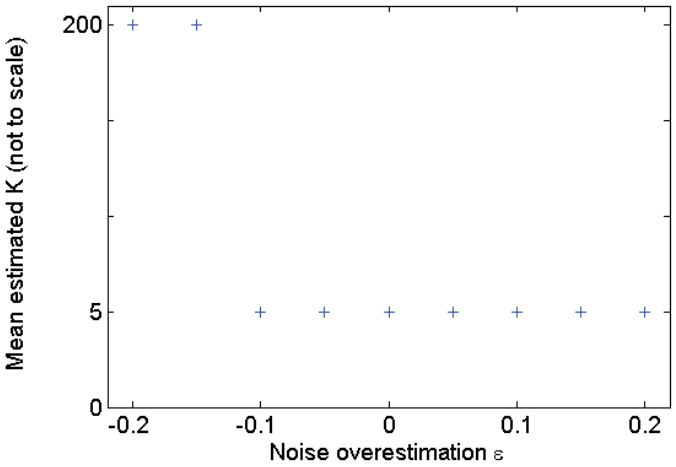


Fig. 3. Mean-estimated ID with regards to the accuracy of the noise estimation. If we assume that the standard deviation of the noise σ is in fact estimated by $\sigma(1+\epsilon)$ (where ϵ is constant), then the results are given for the overestimation, ϵ. Twenty different images were simulated for each value of ϵ to calculate these mean estimates, where the true value for K is 5. Our algorithm only tolerates small underestimations of the noise (up to 10%) and fails for more severe underestimation, but is widely tolerant of overestimation.

underestimated. Even though our method tolerates up to 10% underestimation, it rapidly becomes a limiting factor. In [27] we show that spatially based noise estimation method such as Meer are not sensitive to correlated noise, and that statistical estimations may be improved by a simple band removal process.

V. Results on Real Images

A. Datasets

We test the algorithm on two pairs of real datasets. The first is a pair of hyperspectral images taken over Cuprite, Nevada. One image is an AVIRIS scene, and the other is a SpecTIR scene. The second dataset is a pair of images taken over Lunar Lakes, Nevada, USA. One is an AVIRIS image, and the other is a Hyperion image.

AVIRIS is flown by NASA's Jet Propulsion Lab on the NASA ER-2 aircraft, and is a whiskbroom sensor with approximately 10 nm spectral resolution, over the range 0.4-2.5 μm. This image was obtained from an altitude of 20 km and has a spatial resolution of 20m [28]. The Cuprite image we are considering is a subset from a freely available[1] image taken in 1997, and has 350 $\times$ 350 pixels, 189 bands (bands 1–3, 105–115, and 150–170 were removed from the original data due to water absorption and low SNR, as in [6], [29]), and an average SNR of approximately 500:1 [28]. This was calculated by dividing the mean by the standard deviation of a homogeneous area (Stonewall Playa) and normalising to 50% reflectance [28]. While the exact intrinsic dimension of this dataset is unknown, it is a well-studied image, and Wu *et al.* [6] have tested 6 other methods to determine K in this image. Also, ground truth collected by Swayze *et al.* found at least 18 substances [30] (which may not include rarer minerals), and Chang *et al.* reported that $K = 22$ was the minimum

[1] Available at aviris.jpl.nasa.gov/html/aviris.freedata.html.

number to guarentee that his endmember extraction algorithm could identify five minerals at the ground truth points in the image [29]. The Lunar Lakes scene was acquired in 2009, has 350×350 pixels, and 189 spectral bands.

The SpecTIR subset that we are using contains 320×320 spatial pixels, and 230 spectral bands. SpecTIR is a pushbroom sensor flown commercially. The bands are in the range $0.440 - 1.700$ μm, with approximately 5nm intervals, and the spatial resolution is 1m. The data was processed to reflectance using the same procedures as AVIRIS. The SpecTIR scene is a subset of the AVIRIS scene (but with higher spatial resolution), and so we expect similar results. This dataset is freely available online[2].

Hyperion is a hyperspectral sensor mounted on the EO-1 satellite. Hyperion is capable of producing 220 spectral bands over the spectral range $0.4 - 2.4$ μm, with approximately 10 nm spectral resolution and 30 m spatial resolution. The original images are 7.5 km wide and 100 km long and two separate grating image spectrometers detect VNIR and SWIR wavelengths. Data are available for free online[3]. A subset of the original image is considered, with 250×250 pixels. Hyperion is known to contain very noisy bands and highly correlated bands, and these are removed for this analysis.

The Hyperion and AVIRIS scenes over Lunar Lakes do not cover the exact same spatial area, but they are located in similar areas in the Lunar Lakes region, and so we expect the results over both images to be comparable. The SpecTIR scene over Cuprite is a subset of the AVIRIS scene, but once again we expect the ID estimates to be comparable over both scenes.

B. Results Obtained

The AVIRIS Cuprite image was also tested in [25] and [24]. Constant σ was assumed in [25] and no eigenvalue satisfied the RMT noise evaluation. When whitening was applied in [24], the number of endmembers K, was determined to be 37. This was considered high for the Cuprite area. When using our non-i.i.d. method on the same scene (using Meer's noise approximation method), we calculated that $K = 21$ (as seen in Table I), which is very much in line with the results found in [6]. For the SpecTIR scene our method finds that $K = 24$. We might expect this value to be slightly higher than K determined for AVIRIS, since SpecTIR covers a smaller area at a higher spatial resolution. It is reasonable that at the higher spatial resolution of the SpecTIR sensor, more pure substances should be identifiable.

When analysing Lunar Lakes, Table II shows that the RMT method (again using Meer's noise approximation) found $K = 13$ for AVIRIS and $K = 15$ for Hyperion. In comparison, HFC found ID estimates for Hyperion that are approximately double those of AVIRIS.

C. Statistical Noise Estimation Methods

RMT may be used with any accurate noise approximation method. In this section we test RMT with multiple regression and residual noise approximation methods, and compare the results to HySime [12] (which uses the multiple regression noise estimation) and NSP [8] (which uses the residual noise approximation). The values returned are shown in Table III, where only one value is shown for RMT since both noise approximations yield the same ID estimate. It is obvious that while the AVIRIS results are believable and consistent with the range obtained by HFC and RMT in Table I, the figures for SpecTIR are ridiculous.

The SpecTIR scene over Cuprite contains high levels of spectral correlation—possibly due to the nature of the sensor or to preprocessing of the data. The effects of spectrally correlated noise have been analysed in [27], and these effects explain the ridiculous figures seen in Table III. When the worst of the correlated bands are removed, the results obtained become much more realistic as shown in Table IV. However, only 74 of the original 250 bands remained. In comparison, the AVIRIS scene had only a few badly correlated bands, and so retained all but 8 of its original 189 bands. Nevertheless, the AVIRIS results are also improved. It is unsurprising that now more materials are detected in the AVIRIS scene than in the SpecTIR scene with far fewer spectral bands.

Note that Meer's noise approximation algorithm estimates the noise on a band-by-band basis, and so is not affected by spectrally correlated noise. These results serve as a caution that while RMT may be used with any accurate noise estimation method, statistical noise approximation methods are not accurate in the presence of spectrally correlated noise.

D. Discussion

The algorithm introduced in this paper was able to improve on previous experiments that applied RMT to hyperspectral

TABLE I

APPROXIMATIONS TO K USING HARSANYI-FARRAND-CHANG METHOD (WITH F_D VALUE IN BRACKETS) APPLIED TO AVIRIS AND SPECTIR CUPRITE, FOR DIFFERENT USER-DETERMINED VALUES OF F_D. NOTE THAT THE METHOD IS VERY SENSITIVE TO THESE FALSE ALARM RATES. RMT, USING MEER'S NOISE APPROXIMATION, PRODUCES SIMILAR VALUES TO HFC, BUT THE METHOD DOES NOT REQUIRE ANY PARAMETERS TO BE SET BY THE USER

	HFC(10^{-2})	HFC(10^{-3})	HFC(10^{-4})	RMT
AVIRIS	30	24	22	21
SpecTIR	24	23	19	24

TABLE II

APPROXIMATIONS TO K USING HARSANYI-FARRAND-CHANG METHOD (WITH F_D VALUE IN BRACKETS) APPLIED TO AVIRIS AND HYPERION LUNAR LAKES, FOR DIFFERENT USER-DETERMINED VALUES OF F_D. NOTE THAT THE METHOD IS VERY SENSITIVE TO THESE FALSE ALARM RATES. RMT, USING MEER'S NOISE APPROXIMATION, ESTIMATES THE ID TO BE SIMILAR FOR AVIRIS AND SLIGHTLY SMALLER FOR HYPERION

	HFC(10^{-2})	HFC(10^{-3})	HFC(10^{-4})	RMT
AVIRIS	13	10	9	13
Hyperion	21	20	19	15

[2]Available at www.spectir.com/download.html.

[3]Available at http://edcsns17.cr.usgs.gov/NewEarthExplorer/.

TABLE III
APPROXIMATIONS TO K USING NSP, HYSIME AND RMT WITH STATISTICAL NOISE ESTIMATIONS, APPLIED TO AVIRIS AND SPECTIR CUPRITE

	NSP	RMT	HySime
AVIRIS	28	31	15
SpecTIR	139	156	140

TABLE IV
APPROXIMATIONS TO K USING NSP, HYSIME AND RMT WITH STATISTICAL NOISE ESTIMATIONS, APPLIED TO AVIRIS AND SPECTIR CUPRITE, WITH CORRELATED BANDS REMOVED. HYSIME ESTIMATES THE SAME VALUE FOR BOTH SCENES, AND RMT AND NSP ESTIMATES ARE SIMILAR. NOTE THAT WHEN SPECTRALLY CORRELATED BANDS WERE REMOVED FOR THE STATISTICAL NOISE ESTIMATE, SPECTIR RETAINED LESS THAN HALF THE BANDS RETAINED IN THE AVIRIS IMAGE

	NSP	RMT	HySime
AVIRIS	26	29	17
SpecTIR	21	22	17

imagery. The results were comparable to the well-known method HFC in the Cuprite scene acquired by AVIRIS and SpecTIR (see Table I), with the advantage of not requiring a user-defined parameter, as does HFC. The accuracy of our method may be seen in Tables I, II and IV, where the results of our method are shown to be consistent for both scenes acquired by different sensors.

HFC does not require a noise approximation and so has the advantage of not being dependent on reliable noise estimations. An adaptation of HFC—namely NSP [8]—was developed to reduce sensitivity to the user-defined parameter, although it relies on a statistical noise estimate. Our method is also compared with statistical noise estimates, and when correlated bands are removed, similar values are obtained for the AVIRIS Cuprite scene (see Table IV).

Simulated experiments showed that the RMT method allowed for some error in noise approximation (especially overestimation), although the noise approximation algorithm used should be chosen with care. The method was also accurate up to the same noise levels as HFC, and produced good results for highly variable non-i.i.d. noise.

VI. CONCLUSION

A method has been introduced that uses Random Matrix Theory to determine the intrinsic dimension of a hyperspectral image. This method requires only the assumption of Gaussian noise, (noise may be non-i.i.d. and correlated) and a method of estimating the noise variance in each band. The assumption of Gaussian noise is a common one, and it would be an interesting extension of this work to consider other noise distributions. Our method will accept any noise approximation method, and has been shown to be very tolerant of noise overestimation. In future work, we will test other noise approximation methods in the application to real images. Our method has been successfully tested on real and simulated images, and is comparable to one of the best existing methods, Harsanyi-Farrand-Chang [16], with the advantage that it does not require any user-determined parameters.

APPENDIX

A. Proof of Proposition 1

Proof:

$$\begin{aligned}
\text{Define } S &= \mathbb{E}\left[\underline{\tilde{x}}\underline{\tilde{x}}^T\right] \\
&= \mathbb{E}\Big[(V\underline{\tilde{u}})(V\underline{\tilde{u}})^T + (V\underline{\tilde{u}})\underline{\tilde{\xi}}^T \\
&\qquad + \underline{\tilde{\xi}}(V\underline{\tilde{u}})^T + \underline{\tilde{\xi}}\underline{\tilde{\xi}}^T\Big] \\
&= \mathbb{E}\left[(V\underline{\tilde{u}})(V\underline{\tilde{u}})^T\right] + \mathbb{E}\left[\underline{\tilde{\xi}}\underline{\tilde{\xi}}^T\right] \\
&= V\mathbb{E}\left[\underline{\tilde{u}}\underline{\tilde{u}}^T\right]V^T + \Phi \\
\text{Let } \Pi &= V\mathbb{E}\left[\underline{\tilde{u}}\underline{\tilde{u}}^T\right]V^T \\
\text{Then } S &= \Pi + \Phi
\end{aligned}$$

S is of rank K since $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ is of rank K and V has linearly independent columns. ■

B. Proof of Proposition 2

Proof:

$$\begin{aligned}
\lambda_i^S(\underline{x}_i^\Pi)^T\underline{x}_i^S &= (\underline{x}_i^\Pi)^T S\underline{x}_i^S \\
&= (\underline{x}_i^\Pi)^T(\Pi+\Phi)\underline{x}_i^S \\
&= (\underline{x}_i^\Pi)^T\Pi\underline{x}_i^S + (\underline{x}_i^\Pi)^T\Phi\underline{x}_i^S \\
&= \lambda_i^\Pi(\underline{x}_i^\Pi)^T\underline{x}_i^S + (\underline{x}_i^\Pi)^T\Phi\underline{x}_i^S \\
\lambda_i^S &= \lambda_\Pi + \frac{(\underline{x}_i^\Pi)^T\Phi\underline{x}_i^S}{(\underline{x}_i^\Pi)^T\underline{x}_i^S}
\end{aligned}$$

provided that $(\underline{x}_i^\Pi)^T\underline{x}_i^S \neq 0$. (10) follows at once from the form of Φ. ■

REFERENCES

[1] Z. Ma, "Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Whishart matrices." *Bernoulli*, vol. 18, no. 1, pp. 322–359, 2012.

[2] H.-M. Park, H.-Y. Jung, T.-W. Lee, and S.-Y. Lee, "Subband based blind signal separation for noisy speech recognition," *Electron. Lett.*, vol. 35, no. 23, pp. 2011–2012, Nov. 1999.

[3] F. A. Kruse, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *Int. J. Remote Sensing*, vol. 17, no. 9, pp. 1623–1632, 1996.

[4] D. Gillis, J. Bowles, E. J. Ientilucci, and D. W. Messinger, "A generalised linear mixing model for hyperspectral imagery," *Proc. SPIE*, vol. 6966, pp. 1–11, May 2008.

[5] C.-I. Chang, *Hyperspectal Data Exploitation: Theory and Application*. New York: Wiley, 2007.

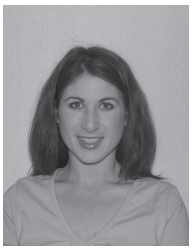
[6] C.-C. Wu, W. Liu, and C.-I. Chang, "Exploration of methods for estimation of number of endmembers in hyperspectral imagery," *Proc. SPIE*, vol. 7, no. 43, pp. 1–11, 2006.

[7] J. Bioucas-Dias and J. Nascimento, "Estimation of signal subspace on hyperspectral data," *Proc. SPIE*, vol. 5982, pp. 191–198, May 2005.

[8] 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.

[9] P. Bajorski, "Does virtual dimensionality work in hyperspectral images?" *Proc. SPIE*, vol. 7334, pp. 1–11, Apr. 2009.

[10] A. Schlamm, D. Messinger, and W. Basener, "Geometric estimation of the inherent dimensionality of single and multi-material clusters in hyperspectral imagery," *J. Appl. Remote Sensing*, vol. 3, no. 1, pp. 033527-1–033527–16, Apr. 2009.
[11] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.
[12] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[13] N. Acito, M. Diani, and G. Corsini, "A new algorithm for robust estimation of the signal subspace in hyperspectral images in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3844–3856, Nov. 2009.
[14] N. Acito, M. Diani, and G. Corsini, "Hyperspectral signal subspace identification in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940–1954, Apr. 2010.
[15] C.-I. Chang, W. Xiong, H.-M. Chen, and J.-W. Chai, "Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 504–520, Jun. 2011.
[16] J. C. Harsanyi, W. Farrand, and C.-I. Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. ASPRS*, 1994, pp. 236–247.
[17] H. Horwitz, R. Nalepka, P. Hyde, and J. Morgenstern, "Estimating the proportions of objects within a single resolution element of a multispectral scanner," in *Proc. 7th Int. Symp. Remote Sensing Environ.*, Jan. 1971, pp. 1307–1320.
[18] P. Meer, J.-M. Jolion, and A. Rosenfeld, "A fast parallel algorithm for blind estimation of noise variance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 216–223, Feb. 1990.
[19] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.
[20] J. Baik and J. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivariate Anal.*, vol. 97, no. 6, pp. 1382–1408, Jul. 2006.
[21] N. El Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Ann. Stat.*, vol. 36, no. 6, pp. 2757–2790, 2008.
[22] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.
[23] B. Nadler and I. M. Johnstone, "Detection performance of Roy's largest root test when the noise covariance is arbitrary," in *Proc. Stat. Signal Process. Workshop*, Jun. 2011, pp. 681–684.
[24] K. Cawse, A. Robin, and M. Sears, "The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image," in *Proc. Workshop Hyperspectral Image Signal Process., Evolut. Remote Sensing*, Jun. 2011, pp. 1–4.
[25] K. Cawse, M. Sears, A. Robin, S. Damelin, F. van den Bergh, K. Wessels, and R. Mathieu, "Using random matrix theory to determine the number of endmembers in a hyperspectral image," in *Proc. Workshop Hyperspectral Image Signal Process., Evol. Remote Sensing*, Jun. 2010, pp. 1–4.
[26] C. I. Grove, S. Hook, and E. Paylor, "Laboratory reflectance spectra of 160 minerals, 0.4 to 2.5 micrometers," JPL Laboratory, NASA, Washington, DC, Tech. Rep. 92–2, 1992.
[27] K. Cawse, A. Robin, and M. Sears, "The effect of spectrally correlated noise on noise estimation methods for hyperspectral images," in *Proc. Workshop Hyperspectral Image Signal Process., Evol., Remote Sensing*, 2012, pp. 1–10.
[28] F. A. Kruse, "Comparison of AVIRIS and hyperion for hyperspectral mineral mapping," in *Proc. 11th JPL Airborne Geosci. Workshop*, 2002, pp. 1–11.
[29] C.-I. Chang, C.-C. Wu, W.-M. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Jul. 2006.
[30] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 47–49.



Kerry Cawse-Nicholson received the Ph.D. degree in computational and applied mathematics from the University of the Witwatersrand, Johannesburg, South Africa, in 2012, while working at the Remote Sensing Research Unit, Meraka Institute, Council for Scientific and Industrial Research, South Africa.

She is currently a Post-Doctoral Researcher with the Rochester Institute of Technology, Rochester, NY, in the field of waveform LiDAR. Her current research interests include image processing, hyperspectral imagery and LiDAR.

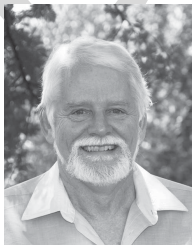
Steven B. Damelin completed his Ph.D. at the The University of the Witwatersrand under the guidance of Doron Lubinsky.

He has held academic positions at the Pennsylvania State University, the University of South Florida, the Institute for Mathematics and its Applications, University of Minnesota, Georgia Southern University, University of the Witwatersrand and at private schools both in South Africa and the United States. Steve has done research in many areas including Computer Vision, Imaging, Signal Processing, Mathematics Education, Learning theory, Computational Harmonic Analysis, Statistical Inference, Harmonic Analysis, Orthogonal Polynomials, Potential Theory, Random Matrices, Approximation Theory, Numerical Analysis, Geometric Analysis and Coding Theory and is a coauthor of a book on signal processing published by Cambridge University Press.

Dr. Damelin was a member of the American Mathematical Society Mathematics Research Communities Advisory Board from February 2011 to January 2014, the distinguished American Mathematical Society Committee of Committees from 2009 to 2010, a New Directions Research Professor at the world renowned Interdisciplinary Institute of Mathematics and its Applications, University of Minnesota. A more detailed bio, CV, and publication list can be found at www.ima.umn.edu/damelin.

Amandine Robin received the Ph.D. degree in applied mathematics from Paris Descartes University, Paris, France, in 2007.

She has been a Lecturer of applied mathematics with the University of the Witwatersrand, Johannesburg, South Africa, since 2009. Her current research interests include statistical modeling for image processing, data fusion and remote sensing, focusing mostly on change detection and classification problems using multi-band remote sensing images (hyperspectral or multi-temporal) or video sequences.



Michael Sears (M'08) received the Ph.D. degree from the Flinders University, Bedford Park, Australia, in 1972.

He was a member of the Department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa, from 1972 until 1997. During this period, he was the Head of the Department, subsequently Dean of the Faculty of Science, and then the Head of the Department of Mathematics. In 1997, he joined Anglo American Corporation as Manager, Remote Sensing and managed hyperspectral sensor surveys and processing. He returned to the University of the Witwatersrand in 2008 and is currently a Visiting Professor with the School of Computer Science. His current research interests include hyperspectral image processing.

Determining the Intrinsic Dimension of a Hyperspectral Image Using Random Matrix Theory

Kerry Cawse-Nicholson, Steven B. Damelin, Amandine Robin, and Michael Sears, *Member, IEEE*

***Abstract*— Determining the intrinsic dimension of a hyperspectral image is an important step in the spectral unmixing process and under- or overestimation of this number may lead to incorrect unmixing in unsupervised methods. In this paper, we discuss a new method for determining the intrinsic dimension using recent advances in random matrix theory. This method is entirely unsupervised, free from any user-determined parameters and allows spectrally correlated noise in the data. Robustness tests are run on synthetic data, to determine how the results were affected by noise levels, noise variability, noise approximation, and spectral characteristics of the endmembers. Success rates are determined for many different synthetic images, and the method is tested on two pairs of real images, namely a Cuprite scene taken from Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) and SpecTIR sensors, and a Lunar Lakes scene taken from AVIRIS and Hyperion, with good results.**

***Index Terms*— Hyperspectral, intrinsic dimension, linear mixture model, random matrix theory, unmixing.**

I. Introduction

DETERMINING the number of sources in a signal is important for the processing of many different types of data, including chemical unmixing [1], extracting speech signals in a noisy line [2], unmixing minerals [3] and unmixing environmental landscapes [4], among many others. Determination of this number is necessary for classification methods, unmixing methods and target detection [5]. It is common practice for the user to select the number of endmembers to suit the application, but this does not necessarily agree with the intrinsic dimension of the image, and an incorrect estimation of this number may have detrimental effects on the end results [6].

Manuscript received July 20, 2011; revised October 24, 2012; accepted October 25, 2012. This work was supported in the part by the Centre for High Performance Computing flagship project: Computational Research Initiative in Imaging and Remote Sensing and CSIR. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian D. Rigling.

K. Cawse-Nicholson is with the School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2000, South Africa, and also with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria 0001, South Africa (e-mail: cawse@cis.rit.edu).

S. B. Damelin is with the Department of Mathematics, Wayne Country Day School, Goldsboro, NC 27530 USA (e-mail: steve.damelin@gmail.com).

A. Robin is with the School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2000, South Africa (e-mail: amandine.c.robin@gmail.com).

M. Sears is with the School of Computer Science, University of the Witwatersrand, Johannesburg 2000, South Africa (e-mail: michael.sears@wits.ac.za).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2227765

In this paper, we will focus on the application to hyperspectral imagery, although this does not preclude applicability to other areas. A hyperspectral image may be visualised as a 3-dimensional data cube of size ($m \times n \times p$). This is a set of p images of size $N = m \times n$ pixels, where all images correspond to the same spatial scene, but are acquired at p different wavelengths.

In a hyperspectral image, the number of bands, p, is large (typically $p \approx 200$). This high spectral resolution enables separation of substances with very similar spectral signatures. However, even with high spatial resolution, each pixel is often a mixture of pure components.

It is of interest to unmix each pixel, to determine the abundances of certain pure substances in the image. In unsupervised remote sensing, the spectra of the K pure substances themselves are also unknown, and the presence of noise forces the dataset to full dimension p. In practice the data set will involve significant redundancy. If we suppose that the data are noise free, then they would be contained in a proper vector subspace $\mathbb{R}^K$ of $\mathbb{R}^p$. Various terms have been introduced in the literature for similar concepts.

Bioucas-Dias and Nascimento [7] define the Intrinsic Dimension of the image (ID) as the dimension of the signal subspace. Chang and Du [8] define ID as the "minimum number of parameters required to account for the observed properties of the data." They also define a separate estimate, called *Virtual Dimension* (VD), which is the number of endmembers necessary to give accurate unmixing, which may be larger than the number of so-called "idealized substances." Here substances may be understood as pure targets or endmembers, and could represent different objects depending on the application, e.g. chemical powders, cover types such as tree species, etc. By definition, VD may be dependent on the unmixing method that is used, unlike ID. Wu *et al.* have compared estimates of VD and ID in their survey paper [6] and found them to be comparable for the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) scene of Cuprite.

Bajorski [9], however, found that the concept of VD may be misleading in certain circumstances, since the value may change when the image is shifted and rotated. Bajorski [9] instead defines the Effective Dimensionality (ED) as "the dimensionality of the affine subspace giving an acceptable approximation to all pixels." Schlamm *et al.* [10] define an Inherent Dimension K', where "the entire spectral image can lie in the same K'-dimensional hyperplane." These authors also claim that the Inherent Dimension is not equivalent to

VD, or the number of spectral signatures present in the image. In contrast, the Spanning Dimension is defined as the "minimum number of basis vectors required to span the space" of pixel observations [10]. Schlamm *et al.* define many other types of dimensions which may apply to hyperspectral imagery, to display the confusion around the term. These authors also define their own intrinsic dimension as "the smallest number of parameters needed to contain all of the variability in the data through a mapping function." This differs from Bioucas-Dias and Nascimento's definition.

The many different definitions of Intrinsic Dimension may be confusing, and although some of the above definitions are closely connected, some definitions may lead to different results, and so it is core for this paper to state a formal definition for ID that is independent of the method used to calculate it.

First, we define some generic notation that will be used in this paper. Underlined variables, such as $\underline{x}$ represent vectors, random variables are denoted by a tilde, $\tilde{x}$, and subscript indices $\underline{x}_i$ refer to measurement $\underline{x}$ for the i^{th} sample.

For each pixel i, $1 \leq i \leq N$, let $\underline{x}_i \in \mathbb{R}^p$ be the observed spectral measurement. Assuming the measurement may be decomposed into signal and noise, write $\underline{x}_i = \underline{s}_i + \underline{\xi}_i$, where $\underline{s}_i$ represents the information in pixel i and $\underline{\xi}_i$ represents the noise. We introduce the following definition.

Definition 1 (Intrinsic Dimension): The Intrinsic Dimension (ID) of a dataset $[\underline{x}_1, \ldots, \underline{x}_N]$ is the dimension, K, of the vector subspace spanned by $\{\underline{s}_i, 1 \leq i \leq N\}$.

Definition 1 is considered to be equivalent to the Bioucas-Dias and Nascimento definition [7], since $\{\underline{s}_1, \ldots, \underline{s}_N\}$ spans the signal subspace, and therefore K is the dimension of the signal subspace. This is also equivalent to the spanning dimension defined in [10].

A. Literature Review

Some of the most popular methods to calculate the ID of a hyperspectral image include Maximum Orthogonal-Complements Algorithm (MOCA) [11], Harsanyi–Farrand–Chang (HFC) [8] and HySime [12]. HySime is a modification of Signal Subspace Estimation (SSE) [7]. MOCA is specifically designed to preserve rare substances and has been improved in terms of computational complexity by Acito *et al.* in [13]. This method has also been combined with HFC by Acito *et al.* [14] and Chang *et al.* [15]. MOCA assumes independent identically distributed (i.i.d.), Gaussian noise, but may be adapted for non-i.i.d. and correlated noise [11].

In 2007, Wu *et al.* [6] provided a summary of methods to determine the intrinsic dimension of a hyperspectral image. This paper compares the HFC Detection Method [16], which assumes white noise with zero mean, with a number of other methods which require various noise assumptions. All the methods discussed only use the eigenvalues of the observation covariance matrix. In all cases, nothing needs to be known about the basis vectors, which is an advantage over supervised methods. These methods are all taken from different areas, including chemistry and signal array processing, and are applied to hyperspectral imagery.

From the real and synthetic experiments described in [6], the authors determined that the best method for hyperspectral imagery was HFC, and the methods with the strictest noise assumptions (i.e. that the noise is Gaussian and i.i.d.) performed the worst. On the other hand, HFC was shown to be sensitive to user-defined values.

B. Linear Mixture Model

A common model used in remote sensing to separate mixed signals is the linear mixture model, introduced by Horwitz *et al.* [17]. This model assumes that the measurement in each pixel is made up of a convex linear combination of pure components. Kritchman and Nadler [1] define the set of pixel observations by the $(p \times N)$ matrix $X = [\underline{x}_1, \ldots, \underline{x}_N]$ and assume that for each pixel indexed by i, the observed measurement vector $\underline{x}_i \in \mathbb{R}^p$ is a realisation of a random vector $\underline{\tilde{x}}$, so that

$$\underline{\tilde{x}} = V\underline{\tilde{u}} + \underline{\tilde{\xi}}, \tag{1}$$

where $\underline{\tilde{u}}$ is a $(K \times 1)$ random vector subject to the constraints $\underline{\tilde{u}} \geq \underline{0}$ and $\underline{1}^T\underline{\tilde{u}} = 1$, and represents the proportions of pure components; V is a $(p \times K)$ deterministic matrix, with columns corresponding to the spectral measurements characterizing the pure components present in the whole scene; $\underline{\tilde{\xi}}$ is a random vector in $\mathbb{R}^p$ representing the noise; and K is the total number of pure components.

It is commonly assumed that $\underline{\tilde{\xi}}$ follows a Gaussian distribution $\mathcal{N}(0, \Phi)$, where Φ is the noise correlation matrix. This model allows the noise to be spectrally but not spatially correlated. This assumption has been successfully used in chemical unmixing [1] and in many methods described in the survey paper by Wu *et al.* [6]. The matrix Φ must be estimated from the data using a noise approximation method such as Meer's method [18].

C. Background on Random Matrix Theory

Determining the ID of a dataset is most often done by analyzing the eigenvalues of the observation covariance matrix, $S(N)$, defined as

$$S(N) := \frac{1}{N}\sum_{i=1}^{N} \underline{x}_i \underline{x}_i^T, \tag{2}$$

where N is the number of samples. Without loss of generality, centered data is assumed, so that the mean pixel value over the entire image is $\underline{0}$. The matrix $S(N)$ is used to distinguish the eigenvalues due to signal and the eigenvalues due to noise. This procedure is used to unmix hyperspectral images [6], signals [2], chemical mixtures [1] and others [3], [4].

In [1], $\underline{\tilde{x}}$ and $\underline{\tilde{\xi}}$ are independent, and $\Phi = \sigma^2 I_p$. This is reasonable in the chemical application being considered, since repeated samples are taken from the same mixture. The model covariance S is defined by $S = \mathbb{E}[\underline{\tilde{x}}\underline{\tilde{x}}^T]$. S may

be diagonalised by W so that

$$W^T S W = W^T V \mathbb{E}\left[\underline{\tilde{u}}\underline{\tilde{u}}^T\right] V^T W + \sigma^2 I_p \tag{3}$$

$$= \begin{pmatrix} l_1 & & 0 \\ & \cdots & \\ & & l_K \\ 0 & & 0 \end{pmatrix} + \sigma^2 I_p \tag{4}$$

where l_i represents the i^{th} signal eigenvalue, and the observed eigenvalues tend toward $l_i + \sigma^2$ for $1 \leq i \leq K$ and to σ^2 for the remaining (noise) eigenvalues. The authors in [1] state that $S(N) \rightarrow S$ as $N \rightarrow \infty$ with probability one. Since in a hyperspectral image N is large, the eigenvalues of the sample covariance might be expected to show this clear distinction between eigenvalues due to signal and eigenvalues due to noise, allowing easy calculation of K. However, in practice, the values of the observed noise eigenvalues can be highly variable, making it difficult to distinguish between a small signal eigenvalue and a large noise eigenvalue. Several methods have been developed to address this problem.

In the method described in [1], the following assumptions are made: the columns of V are linearly independent; the abundance correlation matrix $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ has full rank; and the noise is uncorrelated with the signal. In order to better understand the assumption that $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ has full rank, consider an example where two vectors in V always occur in the same proportions, resulting in reduced rank of $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$. In this case, however, we would expect that the value of K for the scene would be lower than the number of pure substances. Thus this assumption that $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ has rank K is reasonable.

In [1], Kritchman and Nadler have worked with new results in Random Matrix Theory (RMT) to determine which eigenvalues are due to noise and which are due to signal. Because Gaussian noise is assumed, the noise eigenvalues perform like eigenvalues of a random matrix. The distribution of the largest such eigenvalue (the first non-signal eigenvalue) has been well studied [19], [20], and this allows accurate determination of K. The advantage to this method is that there are no parameters that need to be set by the user. It achieved good results in chemical testing.

Wishart matrices play an important part in RMT. Johnstone [19] describes a random cross-product matrix $\tilde{A} = \tilde{X}\tilde{X}^T$, where $\tilde{X}$ contains N independent column vectors, each following a p-variate Gaussian distribution $\mathcal{N}(\mu, \Phi)$. Then $\tilde{A}$ has a p-variate Wishart distribution with N degrees of freedom, $W_p(\Phi, N)$. Johnstone [19] derives results for the case where the mean of each column is zero and the standard deviation of the p-variate Gaussian distribution is the identity matrix, *i.e.* each column of X follows the normal distribution $\mathcal{N}(0, I_p)$, and then $\tilde{A} \sim W_p(I_p, N)$.

Random matrices were first used in physics, to determine quantum energy levels [19]. In this setting, both p and N were large, breaking away from traditional statistics, where p was fixed. And so the authors in [1] assume that $N \rightarrow \infty$, $p \rightarrow \infty$, with $\frac{p}{N} \rightarrow c$, where $c > 0$ is constant. A large body of research in RMT has been dedicated to the distribution of the largest eigenvalue of a matrix following a Wishart distribution [1], [19].

Kritchman and Nadler [1] used some of Johnstone's results [19] in their chemical application to determine the largest noise eigenvalue, by determining the largest sample covariance eigenvalue consistent with the distribution of the largest eigenvalue from a Wishart matrix. This method assumes i.i.d. noise. According to Johnstone [19], the largest eigenvalue of such a real-valued Wishart matrix, $\tilde{\lambda}_1$, fulfills the following condition with convergence in distribution in the limit as $N, p \rightarrow \infty$, $p/N \rightarrow c > 0$:

$$Pr\{\tilde{\lambda}_1 \leq \sigma^2(R_\mu(N, p) + s(\alpha)R_\sigma(N, p))\} \sim TW, \tag{5}$$

where TW denotes the Tracy-Widom distribution, and the rate of convergence is $O(p^{-2/3})$ [1]. In (5), σ^2 is the variance of the Gaussian noise, α is a significance level and $s(\alpha)$ may be found by inverting the Tracy-Widom distribution (in [1], $\alpha = 0.5\%$, and after experimentally investigating the sensitivity to α, we determine that this is optimal and fix this value for all images investigated). The Tracy-Widom distribution is the solution of a second order Painlevé ordinary differential equation [1].

For real valued data define

$$R_\mu(N, p) = \frac{1}{N}\left(\sqrt{N - \frac{1}{2}} + \sqrt{p - \frac{1}{2}}\right)^2 \tag{6}$$

$$R_\sigma(N, p) = \frac{1}{N}\left(\sqrt{N - \frac{1}{2}} + \sqrt{p - \frac{1}{2}}\right) \times \left(\frac{1}{\sqrt{N - \frac{1}{2}}} + \frac{1}{\sqrt{p - \frac{1}{2}}}\right)^{1/3}. \tag{7}$$

Note that these functions do not depend on the intrinsic dimension, K. It is also important that $R_\sigma(N, p) \rightarrow 0$ as p, $N \rightarrow \infty$, $p/N = c$ fixed, so that the $s(\alpha)R_\sigma(N, p)$ term in (5) tends to zero as the image becomes larger. This means that, especially for large images, the formula is not sensitive to the choice of α.

Traditionally, statistical techniques consider the scenario p fixed, with $p << N$, and in this case, the observed noise eigenvalues should be very close to σ^2 in the case of i.i.d. noise [21]. However, although the number of pixels in a hyperspectral image are several orders of magnitude larger than the number of spectral bands, the dividing line between signal and noise eigenvalues is in practice still unclear. El Karoui [21] found that high dimensional problems (a hyperspectral image may contain hundreds of spectral bands) might be better solved by assuming that p and N are both large, with their ratio fixed. Similarly, Kritchman and Nadler [1] state that (5) still holds in the case of finite but large N and p, and in fact their algorithm deals specifically with small N, with good results.

There is, however, a limit to the size of the eigenvalue that may be successfully detected. The phase transition phenomenon described in [20] results in the limit λ_{crit}, below which the noise eigenvalue will not be successfully identified. This limit is defined as

$$\lambda_{crit} = \sigma^2\sqrt{\frac{p}{N}}. \tag{8}$$

The use of RMT in the chemical application proved to be very powerful in the circumstances investigated [1], but the experiments were limited to small numbers of samples, multiple measurements of the same mixture, and i.i.d. noise. This work has been continued by the same authors in [22], where Kritchman and Nadler showed that the method is applicable to a large number of samples in a signal processing environment.

However, the noise was still assumed to be i.i.d. across bands and across samples. Nadler and Johnstone [23] further pursued the idea by considering i.i.d. noise, but possibly correlated samples. They assume the existence of noise only observations and show that the data may be whitened and a variation of Roy's test may be applied.

Since some of the assumptions are not applicable to hyperspectral imagery, in this paper, we extend the RMT ideas from [1] so that they are appropriate for a wider range of applications. We consider large numbers of samples, where each sample may differ greatly from the sample mean, and Section II shows that RMT can be used to develop a new procedure that is applicable to hyperspectral images even in the presence of non-i.i.d. and spectrally correlated noise.

Section III shows the new method in algorithmic format, Section IV shows that this method performs well on synthetic data, and Section V shows that RMT is now applicable to real hyperspectral images. The advantage of the method presented is that it is not dependent on a user-determined parameter, as are other hyperspectral methods discussed above.

II. Model Formulation

Some of the most accurate methods described above have drawbacks in terms of noise assumptions or sensitivity to user-defined parameters. We introduce a method which only assumes Gaussian noise that is independent from the signal (the noise may be correlated across bands, however), and does not rely on user-defined parameters. Specifically, $V\underline{\tilde{u}}$ is independent from $\underline{\tilde{\xi}}$ in (1), and we replace the i.i.d. noise variance $\sigma^2 I_p$ used in [1] with a variance Φ that allows for correlated, non-i.i.d. noise. The core of the method is based on Random Matrix Theory (RMT), and in this section the model will be developed.

First we consider the threshold λ_{crit} in (8), below which noise eigenvalues will not be detected. In hyperspectral images, N is typically much larger than p, and we assume that the eigenvalues we test are above this limit. (Experimentally, in our images this has proven to be true.)

To investigate whether the method described in [1] is applicable to hyperspectral images, we consider a hyperspectral image of pure noise, *i.e.* the measurement $\underline{\tilde{x}}_i$ in each pixel follows a Gaussian distribution $\mathcal{N}(0, \Phi)$. We define $\tilde{X} = [\underline{\tilde{x}}_1, \ldots, \underline{\tilde{x}}_N]$, then as stated above, $\tilde{A} = \tilde{X}\tilde{X}^T$ follows a Wishart distribution $W_p(\Phi, N)$. Denoting $\tilde{S}(N) = N^{-1}\sum_{i=1}^{N} \underline{\tilde{x}}_i \underline{\tilde{x}}_i^T$, which may also be written as $\tilde{S}(N) = N^{-1}\tilde{X}\tilde{X}^T$, then $N\tilde{S}(N)$ follows a Wishart distribution $W_p(\Phi, N)$. Hence, $S(N)$ defined in (2) may be seen as a realisation of $\tilde{S}(N)$ in the case of a pure noise image.

The i.i.d. noise assumption in [1] is not applicable to hyperspectral images, and whitening methods were determined to be unsuccessful in a hyperspectral application [24]. Therefore a new method will be developed to allow for correlated and non-i.i.d. noise.

In [1], the separation between noise and signal eigenvalues is possible because the eigenvalues of S, where $S(N) \to S$ for $N \to \infty$, are given by

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \cdots & \\ & & \lambda_K \\ 0 & & 0 \end{pmatrix} + \sigma^2 I_p.$$

However, this is only true when the standard deviation of the noise is constant across bands. The method using constant noise standard deviation was applied directly to hyperspectral imagery in [25] and, although the method showed good results when applied to synthetic images, on a real image no eigenvalue was found to satisfy (5). This result was to be expected, since it is known that hyperspectral images do not have the same noise standard deviation in each band. Whitening methods were examined in [24], which decorrelate the noise and force the noise standard deviation to become constant in each band so that the original method should be applicable. This improved the previous results, but K was still overestimated. In this paper we will further develop the method so that it is applicable for non-i.i.d. noise, and in order to do that we must first prove that the eigenvalues are still separable in the non-i.i.d. case. To show this we derive Propositions 1 and 2 below.

Proposition 1: Suppose $\underline{\tilde{x}} \in \mathbb{R}^p$ is a random column vector described by $\underline{\tilde{x}} = V\underline{\tilde{u}} + \underline{\tilde{\xi}}$, where V is a $(p \times K)$ matrix with linearly independent columns, $\underline{\tilde{u}}$ is a random vector in $\mathbb{R}^K$, $\underline{\tilde{\xi}}$ is a random vector in $\mathbb{R}^p$, where $\underline{\tilde{\xi}} \sim \mathcal{N}(0, \Phi)$, and Φ is the noise correlation matrix.

Assuming that $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ has full rank K, and that $(V\underline{\tilde{u}})$ and $\underline{\tilde{\xi}}$ are independent, the expectation $S = \mathbb{E}[\underline{\tilde{x}}\underline{\tilde{x}}^T]$ may be decomposed as $S = \Pi + \Phi$, where Π is a symmetric matrix of rank K.

Proof: See Appendix VI-A. ∎

Proposition 1 shows that under an assumption of independence, the signal and noise are separable in the correlation matrix, even if the noise correlation matrix Φ contains varying diagonal terms and/or off-diagonal terms i.e. the noise is not i.i.d.. Random Matrix Theory formulae rely on Φ having the form $\sigma^2 I_p$, so let us write Φ as $\sigma^2 I_p + \bar{\epsilon}$, where σ^2 is the average of the diagonal terms of Φ. Note that S and Φ are symmetric, resulting in symmetric Π. The following proposition proves that the eigenvalues of the observation covariance matrix are also separable, so that the RMT evaluation may be applied.

Proposition 2: Suppose S is a $(p \times p)$ symmetric matrix described by $S = \Pi + \Phi$, where Π and Φ are also symmetric. Suppose $\Phi = \sigma^2 I_p + \bar{\epsilon}$ (σ^2 is scalar). Let λ_i^S be the i^{th} eigenvalue of S (sorted in descending order) and λ_i^Π be the i^{th} eigenvalue of Π (sorted in descending order). Then, assuming

$(\underline{x}_i^{\Pi})^T \underline{x}_i^S \neq 0$

$$\lambda_i^S = \lambda_i^{\Pi} + \frac{(\underline{x}_i^{\Pi})^T \Phi \underline{x}_i^S}{(\underline{x}_i^{\Pi})^T \underline{x}_i^S} \tag{9}$$

$$= \lambda_i^{\Pi} + \sigma^2 + \frac{(\underline{x}_i^{\Pi})^T \bar{\epsilon} \underline{x}_i^S}{(\underline{x}_i^{\Pi})^T \underline{x}_i^S}, \tag{10}$$

where $\underline{x}_i^S$ and $\underline{x}_i^{\Pi}$ are the eigenvectors of S and Π, corresponding to λ_i^S and λ_i^{Π} respectively.

Proof: See Appendix VI-B. ∎

This result is analogous to the one obtained in the case where $\Phi = \sigma^2 I_p$ shown in (4), but with extra terms involving the eigenvectors of Π and S. Propositions 1 and 2 provide a new formula for the noise in the RMT evaluation. For all $1 \leq i \leq p$

$$\rho_i = \frac{(\underline{x}_i^{\Pi})^T \Phi \underline{x}_i^S}{(\underline{x}_i^{\Pi})^T \underline{x}_i^S}, \tag{11}$$

where ρ_i is the difference between the i^{th} observed eigenvalue of S and the i^{th} signal eigenvalue as described by Proposition 2. The original RMT theory in [19] is used to describe the distribution of the largest eigenvalue of a Random Matrix. The authors in [1] have shown that this theory may also be used to detect the largest noise eigenvalue, where the study assumed Gaussian and i.i.d. noise. Since we evaluate each eigenvalue individually, each may be thought of as the largest eigenvalue of a submatrix, where the submatrix of size $(q \times q)$, $q = 1, \ldots, p$, has eigenvalues corresponding to the q smallest eigenvalues of S. So the threshold condition for noise eigenvalues (5) may now be written, for $i = 1, \ldots, p$, as:

$$\lambda_i < \rho_i R,$$
$$\text{where} \quad R = (R_\mu(N, p) + s(\alpha) R_\sigma(N, p)). \tag{12}$$

Now ρ_i depends on knowledge of the noise correlation matrix Φ, the eigenvalues of Π (remember that the observation covariance matrix $S(N)$ approximates $\Pi + \Phi$), and the eigenvalues of $S(N)$. Note that every eigenvalue must be tested, since the value for ρ_i will differ for each evaluation.

In this section the theory has been derived to determine the ID of a hyperspectral image, using RMT. This method allows for non-i.i.d. and correlated noise and requires only an estimate of the noise covariance matrix.

The next section will show how this theory can be used in practice for the estimation of intrinsic dimension, giving a summary of the algorithm used.

III. Algorithm

In section II, we showed how eigenvalue evaluation could be used theoretically to discriminate noise from signal eigenvalues. In order to apply these theoretical results in practice, we need to compute the observation covariance matrix $S(N)$ from the data and approximate the noise correlation matrix Φ.

Note that the RMT significance level, α, may in principle be used as a free parameter as in [22], but we choose to fix this value for all images so that there is no user-determined parameter. This is reasonable because the RMT evaluation is not sensitive to α when images are large, since $s(\alpha)$ is multiplied by $R_\sigma(N, p)$, which tends to zero as $N \to \infty$ (see (5) and (7)). In [1], Kritchman and Nadler fixed alpha to be 0.5%. Using this value in experiments with pure noise images, containing low numbers of pixels (∼1,000) we found that 99.7% of the time all eigenvalues were ascribed to noise. On the other hand, if a single signal is included in the image, one eigenvalue was always ascribed to signal. From these experiments this is indeed a reasonable value for α for hyperspectral data and everywhere in this paper it will be assumed that $\alpha = 0.5\%$. It should be noted that α is not dependent on the image (as it is fixed for all images) and is not considered a tuneable parameter.

Algorithm 1

$S(N) \leftarrow$ observation covariance matrix
$\underline{\lambda} \leftarrow [\lambda_1, \ldots, \lambda_p]$, the sorted (descending) eigenvalues of $S(N)$
$p \leftarrow \text{length}(\underline{\lambda})$
$R_\mu \leftarrow \frac{1}{N}(\sqrt{N - 0.5} + \sqrt{p - 0.5})^2$
$R_\sigma \leftarrow \frac{1}{N}(\sqrt{N - 0.5} + \sqrt{p - 0.5})(\frac{1}{\sqrt{N-0.5}} + \frac{1}{\sqrt{p-0.5}})^{\frac{1}{3}}$
$s \leftarrow (-\frac{3}{2}\log[4\sqrt{\pi}\frac{\alpha}{100}])^{\frac{2}{3}}$
$R \leftarrow R_\mu + s \times R_\sigma$
$f \leftarrow$ call NoiseApproximationFunction(Image)
$\Phi \leftarrow$ diagonal matrix containing variances per band determined by f
$\sigma^2 \leftarrow$ mean of the noise variances over all bands *i.e.* mean($[\Phi_{1,1}, \ldots, \Phi_{p,p}]$)
$E_1 \leftarrow$ eigenvectors of $S(N)$ (sorted so that corresponding eigenvalues are descending)
$E_2 \leftarrow$ eigenvectors of $(S(N) - \Phi)$ (sorted so that corresponding eigenvalues are descending)
for $i = 1 \rightarrow p$ **do**
  $\rho_i \leftarrow \frac{E_1^i . \Phi . E_2^i}{E_1^i . E_2^i}$
end for
$\underline{\rho} \leftarrow [\rho_1, \cdots, \rho_p]$
$K \leftarrow$ intersection between $\underline{\lambda}$ and $\underline{\rho} \times R$

To approximate the noise, we chose an algorithm described by Meer *et al.* in [18]. This is a pyramid algorithm that searches for homogeneous areas, and it has various rules to allow for outliers. In test images, this algorithm was able to approximate the noise accurately with RMSE of 1%. This particular algorithm produces only the variance per band, so in this case Φ is diagonal. Since we are assuming this noise approximation method, from this point on we assume a diagonal noise covariance matrix. Note that this is not a restriction to our method, any preferred noise approximation algorithm may be used. Meer's method was chosen since it estimates the noise on a band-by-band basis, and thus it is independent of any relationship between bands, such as noise correlation. The sensitivity of our method with respect to the accuracy of the noise approximation will be tested and discussed in Section IV-B.

Unlike previous versions [1], [25], our method allows for non-i.i.d. noise. This procedure is summarised in algorithmic format above.

The algorithm requires only the image itself as an input, and returns the intrinsic dimension of the image, K. It requires no prior knowledge about the image, and requires no user-determined inputs. In the next section, the robustness of this approach is analysed with regards to each variable of the model, using simulated data.

IV. Performance Analysis

A. Description of the Test Dataset

We simulated a dataset using 20 minerals taken from the JPL spectral library [26]. At each iteration of the algorithm, K unique minerals are chosen randomly from the dataset, and the proportions of each of the basis vectors in each pixel are randomly selected with the only restrictions being the positive and sum-to-one conditions that are enforced. We also varied our values for p (number of bands), N (number of pixels), and σ (the average of noise standard deviations across all bands) in these images. As seen in Figure 1, the spectra in our dataset vary widely—some are similar and others easily separable. By mixing random spectra together we simulate images that may be easy or difficult to process. Since the proportions are randomly selected as well, one or more basis vector may occur in very small proportions in the image. All of these properties were created in order to simulate as realistic a dataset as possible.

While pure endmembers may be well understood in mineral applications, this idea becomes more subjective in data that contains vegetation spectra, which may vary according to species, phenology, etc. In this paper it is assumed that each target may be represented by a single spectral signature, where variation around this target is understood to be smaller in magnitude than the noise present in the image. If the variation is larger than this, then certain vegetation classes may be represented by more than one spectral signature.

In each test, we fixed all variables (except the one being examined) at the following default values: $N = 10,000$, $p = 200$, $\sigma = 10^{-3}$, $K = 5$.

When noise statistics are assumed known, the method is independent of N and p. In this case, the method also produces 100% success rates for non-i.i.d. noise and noise that is correlated between bands. Note that in all tests below, the actual noise values used to construct the synthetic dataset are known, so that we test the robustness of the ID method rather than the accuracy of any noise approximation method. If the noise is approximated, accuracy will decrease.

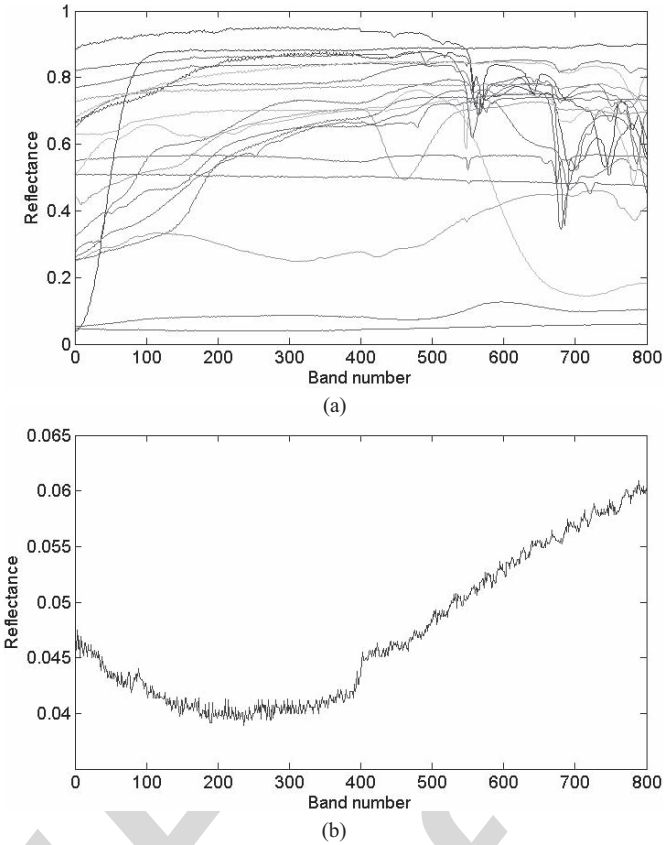


Fig. 1. (a) All the spectra used to make up the test data set. Note that there is a mix between similar and easily separable spectra, flat spectra and those with sharp features, spectra with high and low amplitude, etc. This is done to mimic a real environment. (b) Spectrum (Graphite 1A) with the lowest amplitude is shown at a larger scale. Take note of the noisiness of the spectrum.

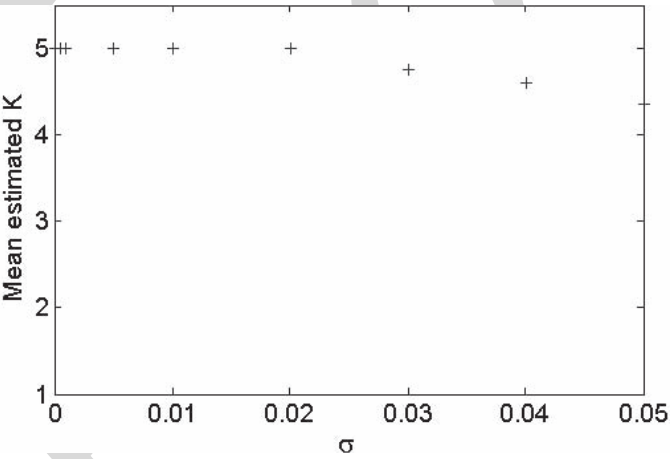


Fig. 2. Mean-estimated ID with regards to different noise levels. Twenty different images per noise level were simulated to calculate these values, and the true value for K is 5. Note that the method produces the correct value $K = 5$ in all 20 simulations for noise standard deviation up to 0.02.

B. Robustness to the Noise Level

Noise is an important element to test, as high noise eigenvalues can become very close to small signal values, making them difficult to separate. With low SNR (signal to noise ratio), two similar spectra will become indistinguishable. As seen in Figure 1, our test dataset does contain some basis vectors which are very close together, and so we are interested to test the level at which they are impossible to separate. Our tests in Figure 2 show that the mean estimated K, over 20 simulations, is exactly 5 for noise standard deviation below $\sigma = 0.02$. We use the SNR definition in [6] which says that the SNR is equal to half the mean signal divided by the standard deviation of the noise. Since our synthetic image mean is close to 0.5, a noise value of $\sigma = 0.01$ is equivalent to an SNR of 25:1, which is the limit of the noise tested in [8]. This lower bound is much lower than current hyperspectral sensors, with AVIRIS having an SNR of approximately 500:1 [3]. It should

be noted, however, that the intrinsic dimension will not be accurately determined for images containing very high levels of noise.

Note in Figure 1(a) that there are two spectra of a much lower amplitude than the rest. We will discuss these in Section IV-C, and study the impact of the number of basis vectors on the performance.

C. Performance With Regards to the Spectral Characteristics of Basis Vectors

Certain characteristics of basis vectors will make them difficult to detect in any model. Multiple spectra that have similar spectral signatures may be difficult to separate, and spectra with low amplitudes may be confused with noise. We have investigated the impact of such scenarios in our simulated dataset. Figure 1 clearly shows that our dataset contains spectra with different characteristics in order to mimic a real environment. Some spectra may appear to be similar, but our method was able to separate these. Two spectra stand out with very low amplitudes, and one spectrum in particular, Graphite, (Figure 1(b)) has low amplitude and high noise (mean reflectance 0.0475 and standard deviation 0.007). This spectrum was impossible to distinguish from noise with the default noise standard deviation of $\sigma = 10^{-3}$. This is equivalent to an SNR of 250:1, which is still low compared to real images (recall that AVIRIS has SNR = 500:1). Note that if the noise is reduced to $\sigma = 10^{-4}$, the basis vectors are separated with 100% accuracy. It is interesting that the other low-amplitude spectrum is still accurately detected—it may be the combination of the noisy signal and the low amplitude in Graphite that causes it to be recognised as noise in a noisy image.

D. Robustness to Noise Approximation

As previously stated, our algorithm does not depend on a specific noise approximation algorithm. However, the algorithm that is used must be accurate enough for our purposes. In this section, we test the impact of errors in noise approximation on the accuracy of our method in order to evaluate how accurate the noise approximation needs to be.

If we assume that the noise distribution in all bands is $\mathcal{N}(0, \sigma^2 I_p)$, Figure 3 shows that our algorithm only tolerates up to 10% underestimation, and overestimations of constant noise are well tolerated. This asymmetrical behaviour is firstly due to the eigenvalue spread (noise eigenvalues are closer together, so a noise approximation error towards the noise direction will result in higher errors in estimating K), and secondly due to the fact that the RMT evaluation in (5) is a single sided inequality; i.e. if the upper limit for the largest noise eigenvalue is increased, the noise eigenvalue will still fall below this limit. Conversely, if it is decreased, some noise eigenvalues will be recognised as signal. Again, this is made more predominant due to the asymmetrical spread of the eigenvalues. Therefore, for the best results, the noise approximation algorithm should be chosen with care, since the ID may not be accurate if noise is underestimated. Note that when the noise is spectrally correlated, it is often

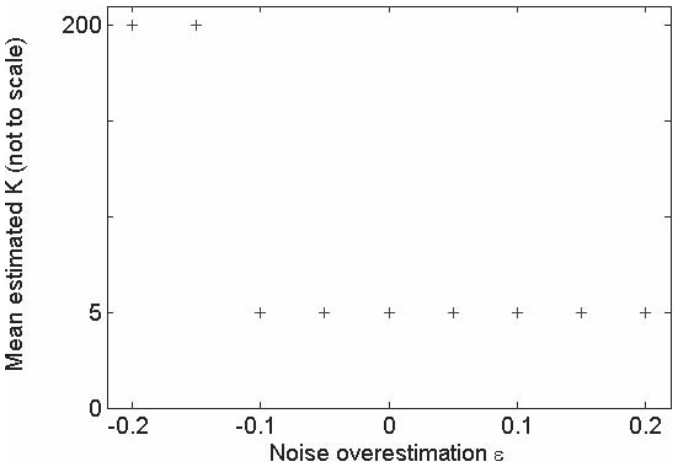


Fig. 3. Mean-estimated ID with regards to the accuracy of the noise estimation. If we assume that the standard deviation of the noise σ is in fact estimated by $\sigma(1+\epsilon)$ (where ϵ is constant), then the results are given for the overestimation, ϵ. Twenty different images were simulated for each value of ϵ to calculate these mean estimates, where the true value for K is 5. Our algorithm only tolerates small underestimations of the noise (up to 10%) and fails for more severe underestimation, but is widely tolerant of overestimation.

underestimated. Even though our method tolerates up to 10% underestimation, it rapidly becomes a limiting factor. In [27] we show that spatially based noise estimation method such as Meer are not sensitive to correlated noise, and that statistical estimations may be improved by a simple band removal process.

V. Results on Real Images

A. Datasets

We test the algorithm on two pairs of real datasets. The first is a pair of hyperspectral images taken over Cuprite, Nevada. One image is an AVIRIS scene, and the other is a SpecTIR scene. The second dataset is a pair of images taken over Lunar Lakes, Nevada, USA. One is an AVIRIS image, and the other is a Hyperion image.

AVIRIS is flown by NASA's Jet Propulsion Lab on the NASA ER-2 aircraft, and is a whiskbroom sensor with approximately 10 nm spectral resolution, over the range 0.4-2.5 μm. This image was obtained from an altitude of 20 km and has a spatial resolution of 20m [28]. The Cuprite image we are considering is a subset from a freely available[1] image taken in 1997, and has 350 × 350 pixels, 189 bands (bands 1–3, 105–115, and 150–170 were removed from the original data due to water absorption and low SNR, as in [6], [29]), and an average SNR of approximately 500:1 [28]. This was calculated by dividing the mean by the standard deviation of a homogeneous area (Stonewall Playa) and normalising to 50% reflectance [28]. While the exact intrinsic dimension of this dataset is unknown, it is a well-studied image, and Wu *et al.* [6] have tested 6 other methods to determine K in this image. Also, ground truth collected by Swayze *et al.* found at least 18 substances [30] (which may not include rarer minerals), and Chang *et al.* reported that $K = 22$ was the minimum

[1]Available at aviris.jpl.nasa.gov/html/aviris.freedata.html.

number to guarentee that his endmember extraction algorithm could identify five minerals at the ground truth points in the image [29]. The Lunar Lakes scene was acquired in 2009, has 350×350 pixels, and 189 spectral bands.

The SpecTIR subset that we are using contains 320×320 spatial pixels, and 230 spectral bands. SpecTIR is a pushbroom sensor flown commercially. The bands are in the range $0.440 - 1.700$ μm, with approximately 5nm intervals, and the spatial resolution is 1m. The data was processed to reflectance using the same procedures as AVIRIS. The SpecTIR scene is a subset of the AVIRIS scene (but with higher spatial resolution), and so we expect similar results. This dataset is freely available online[2].

Hyperion is a hyperspectral sensor mounted on the EO-1 satellite. Hyperion is capable of producing 220 spectral bands over the spectral range $0.4 - 2.4$ μm, with approximately 10 nm spectral resolution and 30 m spatial resolution. The original images are 7.5 km wide and 100 km long and two separate grating image spectrometers detect VNIR and SWIR wavelengths. Data are available for free online[3]. A subset of the original image is considered, with 250×250 pixels. Hyperion is known to contain very noisy bands and highly correlated bands, and these are removed for this analysis.

The Hyperion and AVIRIS scenes over Lunar Lakes do not cover the exact same spatial area, but they are located in similar areas in the Lunar Lakes region, and so we expect the results over both images to be comparable. The SpecTIR scene over Cuprite is a subset of the AVIRIS scene, but once again we expect the ID estimates to be comparable over both scenes.

B. Results Obtained

The AVIRIS Cuprite image was also tested in [25] and [24]. Constant σ was assumed in [25] and no eigenvalue satisfied the RMT noise evaluation. When whitening was applied in [24], the number of endmembers K, was determined to be 37. This was considered high for the Cuprite area. When using our non-i.i.d. method on the same scene (using Meer's noise approximation method), we calculated that $K = 21$ (as seen in Table I), which is very much in line with the results found in [6]. For the SpecTIR scene our method finds that $K = 24$. We might expect this value to be slightly higher than K determined for AVIRIS, since SpecTIR covers a smaller area at a higher spatial resolution. It is reasonable that at the higher spatial resolution of the SpecTIR sensor, more pure substances should be identifiable.

When analysing Lunar Lakes, Table II shows that the RMT method (again using Meer's noise approximation) found $K = 13$ for AVIRIS and $K = 15$ for Hyperion. In comparison, HFC found ID estimates for Hyperion that are approximately double those of AVIRIS.

C. Statistical Noise Estimation Methods

RMT may be used with any accurate noise approximation method. In this section we test RMT with multiple regression and residual noise approximation methods, and compare the results to HySime [12] (which uses the multiple regression noise estimation) and NSP [8] (which uses the residual noise approximation). The values returned are shown in Table III, where only one value is shown for RMT since both noise approximations yield the same ID estimate. It is obvious that while the AVIRIS results are believable and consistent with the range obtained by HFC and RMT in Table I, the figures for SpecTIR are ridiculous.

The SpecTIR scene over Cuprite contains high levels of spectral correlation—possibly due to the nature of the sensor or to preprocessing of the data. The effects of spectrally correlated noise have been analysed in [27], and these effects explain the ridiculous figures seen in Table III. When the worst of the correlated bands are removed, the results obtained become much more realistic as shown in Table IV. However, only 74 of the original 250 bands remained. In comparison, the AVIRIS scene had only a few badly correlated bands, and so retained all but 8 of its original 189 bands. Nevertheless, the AVIRIS results are also improved. It is unsurprising that now more materials are detected in the AVIRIS scene than in the SpecTIR scene with far fewer spectral bands.

Note that Meer's noise approximation algorithm estimates the noise on a band-by-band basis, and so is not affected by spectrally correlated noise. These results serve as a caution that while RMT may be used with any accurate noise estimation method, statistical noise approximation methods are not accurate in the presence of spectrally correlated noise.

D. Discussion

The algorithm introduced in this paper was able to improve on previous experiments that applied RMT to hyperspectral

TABLE I

APPROXIMATIONS TO K USING HARSANYI-FARRAND-CHANG METHOD (WITH F_D VALUE IN BRACKETS) APPLIED TO AVIRIS AND SPECTIR CUPRITE, FOR DIFFERENT USER-DETERMINED VALUES OF F_D. NOTE THAT THE METHOD IS VERY SENSITIVE TO THESE FALSE ALARM RATES. RMT, USING MEER'S NOISE APPROXIMATION, PRODUCES SIMILAR VALUES TO HFC, BUT THE METHOD DOES NOT REQUIRE ANY PARAMETERS TO BE SET BY THE USER

	HFC(10^{-2})	HFC(10^{-3})	HFC(10^{-4})	RMT
AVIRIS	30	24	22	21
SpecTIR	24	23	19	24

TABLE II

APPROXIMATIONS TO K USING HARSANYI-FARRAND-CHANG METHOD (WITH F_D VALUE IN BRACKETS) APPLIED TO AVIRIS AND HYPERION LUNAR LAKES, FOR DIFFERENT USER-DETERMINED VALUES OF F_D. NOTE THAT THE METHOD IS VERY SENSITIVE TO THESE FALSE ALARM RATES. RMT, USING MEER'S NOISE APPROXIMATION, ESTIMATES THE ID TO BE SIMILAR FOR AVIRIS AND SLIGHTLY SMALLER FOR HYPERION

	HFC(10^{-2})	HFC(10^{-3})	HFC(10^{-4})	RMT
AVIRIS	13	10	9	13
Hyperion	21	20	19	15

[2]Available at www.spectir.com/download.html.

[3]Available at http://edcsns17.cr.usgs.gov/NewEarthExplorer/.

TABLE III
APPROXIMATIONS TO K USING NSP, HYSIME AND RMT WITH STATISTICAL NOISE ESTIMATIONS, APPLIED TO AVIRIS AND SPECTIR CUPRITE

	NSP	RMT	HySime
AVIRIS	28	31	15
SpecTIR	139	156	140

TABLE IV
APPROXIMATIONS TO K USING NSP, HYSIME AND RMT WITH STATISTICAL NOISE ESTIMATIONS, APPLIED TO AVIRIS AND SPECTIR CUPRITE, WITH CORRELATED BANDS REMOVED. HYSIME ESTIMATES THE SAME VALUE FOR BOTH SCENES, AND RMT AND NSP ESTIMATES ARE SIMILAR. NOTE THAT WHEN SPECTRALLY CORRELATED BANDS WERE REMOVED FOR THE STATISTICAL NOISE ESTIMATE, SPECTIR RETAINED LESS THAN HALF THE BANDS RETAINED IN THE AVIRIS IMAGE

	NSP	RMT	HySime
AVIRIS	26	29	17
SpecTIR	21	22	17

imagery. The results were comparable to the well-known method HFC in the Cuprite scene acquired by AVIRIS and SpecTIR (see Table I), with the advantage of not requiring a user-defined parameter, as does HFC. The accuracy of our method may be seen in Tables I, II and IV, where the results of our method are shown to be consistent for both scenes acquired by different sensors.

HFC does not require a noise approximation and so has the advantage of not being dependent on reliable noise estimations. An adaptation of HFC—namely NSP [8]—was developed to reduce sensitivity to the user-defined parameter, although it relies on a statistical noise estimate. Our method is also compared with statistical noise estimates, and when correlated bands are removed, similar values are obtained for the AVIRIS Cuprite scene (see Table IV).

Simulated experiments showed that the RMT method allowed for some error in noise approximation (especially overestimation), although the noise approximation algorithm used should be chosen with care. The method was also accurate up to the same noise levels as HFC, and produced good results for highly variable non-i.i.d. noise.

VI. CONCLUSION

A method has been introduced that uses Random Matrix Theory to determine the intrinsic dimension of a hyperspectral image. This method requires only the assumption of Gaussian noise, (noise may be non-i.i.d. and correlated) and a method of estimating the noise variance in each band. The assumption of Gaussian noise is a common one, and it would be an interesting extension of this work to consider other noise distributions. Our method will accept any noise approximation method, and has been shown to be very tolerant of noise overestimation. In future work, we will test other noise approximation methods in the application to real images. Our method has been successfully tested on real and simulated images, and is comparable to one of the best existing methods, Harsanyi-Farrand-Chang [16], with the advantage that it does not require any user-determined parameters.

APPENDIX

A. Proof of Proposition 1

Proof:

$$\begin{aligned}
\text{Define } S &= \mathbb{E}\left[\underline{\tilde{x}}\underline{\tilde{x}}^T\right] \\
&= \mathbb{E}\Big[(V\underline{\tilde{u}})(V\underline{\tilde{u}})^T + (V\underline{\tilde{u}})\underline{\tilde{\xi}}^T \\
&\qquad + \underline{\tilde{\xi}}(V\underline{\tilde{u}})^T + \underline{\tilde{\xi}}\underline{\tilde{\xi}}^T\Big] \\
&= \mathbb{E}\left[(V\underline{\tilde{u}})(V\underline{\tilde{u}})^T\right] + \mathbb{E}\left[\underline{\tilde{\xi}}\underline{\tilde{\xi}}^T\right] \\
&= V\mathbb{E}\left[\underline{\tilde{u}}\underline{\tilde{u}}^T\right]V^T + \Phi \\
\text{Let } \Pi &= V\mathbb{E}\left[\underline{\tilde{u}}\underline{\tilde{u}}^T\right]V^T \\
\text{Then } S &= \Pi + \Phi
\end{aligned}$$

S is of rank K since $\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]$ is of rank K and V has linearly independent columns. ■

B. Proof of Proposition 2

Proof:

$$\begin{aligned}
\lambda_i^S(\underline{x}_i^\Pi)^T\underline{x}_i^S &= (\underline{x}_i^\Pi)^T S\underline{x}_i^S \\
&= (\underline{x}_i^\Pi)^T(\Pi+\Phi)\underline{x}_i^S \\
&= (\underline{x}_i^\Pi)^T\Pi\underline{x}_i^S + (\underline{x}_i^\Pi)^T\Phi\underline{x}_i^S \\
&= \lambda_i^\Pi(\underline{x}_i^\Pi)^T\underline{x}_i^S + (\underline{x}_i^\Pi)^T\Phi\underline{x}_i^S \\
\lambda_i^S &= \lambda_\Pi + \frac{(\underline{x}_i^\Pi)^T\Phi\underline{x}_i^S}{(\underline{x}_i^\Pi)^T\underline{x}_i^S}
\end{aligned}$$

provided that $(\underline{x}_i^\Pi)^T\underline{x}_i^S \neq 0$. (10) follows at once from the form of Φ. ■

REFERENCES

[1] Z. Ma, "Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Whishart matrices." *Bernoulli*, vol. 18, no. 1, pp. 322–359, 2012.

[2] H.-M. Park, H.-Y. Jung, T.-W. Lee, and S.-Y. Lee, "Subband based blind signal separation for noisy speech recognition," *Electron. Lett.*, vol. 35, no. 23, pp. 2011–2012, Nov. 1999.

[3] F. A. Kruse, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *Int. J. Remote Sensing*, vol. 17, no. 9, pp. 1623–1632, 1996.

[4] D. Gillis, J. Bowles, E. J. Ientilucci, and D. W. Messinger, "A generalised linear mixing model for hyperspectral imagery," *Proc. SPIE*, vol. 6966, pp. 1–11, May 2008.

[5] C.-I. Chang, *Hyperspectal Data Exploitation: Theory and Application*. New York: Wiley, 2007.

[6] C.-C. Wu, W. Liu, and C.-I. Chang, "Exploration of methods for estimation of number of endmembers in hyperspectral imagery," *Proc. SPIE*, vol. 7, no. 43, pp. 1–11, 2006.

[7] J. Bioucas-Dias and J. Nascimento, "Estimation of signal subspace on hyperspectral data," *Proc. SPIE*, vol. 5982, pp. 191–198, May 2005.

[8] 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.

[9] P. Bajorski, "Does virtual dimensionality work in hyperspectral images?" *Proc. SPIE*, vol. 7334, pp. 1–11, Apr. 2009.

[10] A. Schlamm, D. Messinger, and W. Basener, "Geometric estimation of the inherent dimensionality of single and multi-material clusters in hyperspectral imagery," *J. Appl. Remote Sensing*, vol. 3, no. 1, pp. 033527-1–033527–16, Apr. 2009.
[11] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.
[12] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[13] N. Acito, M. Diani, and G. Corsini, "A new algorithm for robust estimation of the signal subspace in hyperspectral images in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3844–3856, Nov. 2009.
[14] N. Acito, M. Diani, and G. Corsini, "Hyperspectral signal subspace identification in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940–1954, Apr. 2010.
[15] C.-I. Chang, W. Xiong, H.-M. Chen, and J.-W. Chai, "Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 504–520, Jun. 2011.
[16] J. C. Harsanyi, W. Farrand, and C.-I. Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. ASPRS*, 1994, pp. 236–247.
[17] H. Horwitz, R. Nalepka, P. Hyde, and J. Morgenstern, "Estimating the proportions of objects within a single resolution element of a multispectral scanner," in *Proc. 7th Int. Symp. Remote Sensing Environ.*, Jan. 1971, pp. 1307–1320.
[18] P. Meer, J.-M. Jolion, and A. Rosenfeld, "A fast parallel algorithm for blind estimation of noise variance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 216–223, Feb. 1990.
[19] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.
[20] J. Baik and J. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivariate Anal.*, vol. 97, no. 6, pp. 1382–1408, Jul. 2006.
[21] N. El Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Ann. Stat.*, vol. 36, no. 6, pp. 2757–2790, 2008.
[22] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.
[23] B. Nadler and I. M. Johnstone, "Detection performance of Roy's largest root test when the noise covariance is arbitrary," in *Proc. Stat. Signal Process. Workshop*, Jun. 2011, pp. 681–684.
[24] K. Cawse, A. Robin, and M. Sears, "The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image," in *Proc. Workshop Hyperspectral Image Signal Process., Evolut. Remote Sensing*, Jun. 2011, pp. 1–4.
[25] K. Cawse, M. Sears, A. Robin, S. Damelin, F. van den Bergh, K. Wessels, and R. Mathieu, "Using random matrix theory to determine the number of endmembers in a hyperspectral image," in *Proc. Workshop Hyperspectral Image Signal Process., Evol. Remote Sensing*, Jun. 2010, pp. 1–4.
[26] C. I. Grove, S. Hook, and E. Paylor, "Laboratory reflectance spectra of 160 minerals, 0.4 to 2.5 micrometers," JPL Laboratory, NASA, Washington, DC, Tech. Rep. 92–2, 1992.
[27] K. Cawse, A. Robin, and M. Sears, "The effect of spectrally correlated noise on noise estimation methods for hyperspectral images," in *Proc. Workshop Hyperspectral Image Signal Process., Evol., Remote Sensing*, 2012, pp. 1–10.
[28] F. A. Kruse, "Comparison of AVIRIS and hyperion for hyperspectral mineral mapping," in *Proc. 11th JPL Airborne Geosci. Workshop*, 2002, pp. 1–11.
[29] C.-I. Chang, C.-C. Wu, W.-M. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Jul. 2006.
[30] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 47–49.

Kerry Cawse-Nicholson received the Ph.D. degree in computational and applied mathematics from the University of the Witwatersrand, Johannesburg, South Africa, in 2012, while working at the Remote Sensing Research Unit, Meraka Institute, Council for Scientific and Industrial Research, South Africa.

She is currently a Post-Doctoral Researcher with the Rochester Institute of Technology, Rochester, NY, in the field of waveform LiDAR. Her current research interests include image processing, hyperspectral imagery and LiDAR.

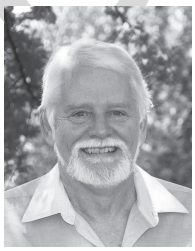
Steven B. Damelin completed his Ph.D. at the The University of the Witwatersrand under the guidance of Doron Lubinsky.

He has held academic positions at the Pennsylvania State University, the University of South Florida, the Institute for Mathematics and its Applications, University of Minnesota, Georgia Southern University, University of the Witwatersrand and at private schools both in South Africa and the United States. Steve has done research in many areas including Computer Vision, Imaging, Signal Processing, Mathematics Education, Learning theory, Computational Harmonic Analysis, Statistical Inference, Harmonic Analysis, Orthogonal Polynomials, Potential Theory, Random Matrices, Approximation Theory, Numerical Analysis, Geometric Analysis and Coding Theory and is a coauthor of a book on signal processing published by Cambridge University Press.

Dr. Damelin was a member of the American Mathematical Society Mathematics Research Communities Advisory Board from February 2011 to January 2014, the distinguished American Mathematical Society Committee of Committees from 2009 to 2010, a New Directions Research Professor at the world renowned Interdisciplinary Institute of Mathematics and its Applications, University of Minnesota. A more detailed bio, CV, and publication list can be found at www.ima.umn.edu/damelin.

Amandine Robin received the Ph.D. degree in applied mathematics from Paris Descartes University, Paris, France, in 2007.

She has been a Lecturer of applied mathematics with the University of the Witwatersrand, Johannesburg, South Africa, since 2009. Her current research interests include statistical modeling for image processing, data fusion and remote sensing, focusing mostly on change detection and classification problems using multi-band remote sensing images (hyperspectral or multi-temporal) or video sequences.



Michael Sears (M'08) received the Ph.D. degree from the Flinders University, Bedford Park, Australia, in 1972.

He was a member of the Department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa, from 1972 until 1997. During this period, he was the Head of the Department, subsequently Dean of the Faculty of Science, and then the Head of the Department of Mathematics. In 1997, he joined Anglo American Corporation as Manager, Remote Sensing and managed hyperspectral sensor surveys and processing. He returned to the University of the Witwatersrand in 2008 and is currently a Visiting Professor with the School of Computer Science. His current research interests include hyperspectral image processing.