THE EFFECT OF NOISE WHITENING ON METHODS FOR DETERMINING THE INTRINSIC DIMENSION OF A HYPERSPECTRAL IMAGE

K. Cawse[1,2]*, A. Robin*[1]*, M. Sears*[3]

[1]School of Computational and Applied Maths
[3]School of Computer Science
University of the Witwatersrand
South Africa

[2]Remote Sensing Research Unit
Meraka Institute, CSIR
South Africa

ABSTRACT

Determining the intrinsic dimension of a hyperspectral image is an important step in the spectral unmixing process, and under- or over- estimation of this number may lead to incorrect unmixing for unsupervised methods. It is known that most real images contain noise that is not i.i.d. across bands, and so methods that assume i.i.d. noise are often avoided. However, this problem may be alleviated by implementing a noise whitening procedure as a pre-processing step. In this paper we will investigate one particular noise whitening approach, as well as a noise removal approach, and consider how the application of these methods may improve several methods for determining the intrinsic dimension of an image, including Malinowski's Empirical Indicator Function [1], Random Matrix Theory [2], and Harsanyi-Farrand-Chang [3].

Index Terms— Hyperspectral Unmixing, Random Matrix Theory, Intrinsic Dimension, Noise Whitening.

1. INTRODUCTION

Determining the intrinsic dimension (the dimension of the signal subspace) of an image is important for the processing of many different types of data, including chemical unmixing, extracting speech signals, unmixing minerals and unmixing environmental landscapes, among many others.

A common model used to unmix hyperspectral images is the linear mixing model. This model assumes that each pixel in the image is made up of a linear combination of "endmembers". Mathematically, consider the reflectance measured in each pixel i as a vector $\mathbf{x}_i = [x_{i1}, \ldots, x_{ip}]^T$ for $1 \leq i \leq N$. The linear mixture model states that, for each pixel i,

$$\mathbf{x}_i = \sum_{j=1}^{K} a_{ij}\mathbf{v}_j + \mathbf{n}_i \tag{1}$$

where a_{ij} represents the proportion of endmember $\mathbf{v}_j$ in the mixed pixel i, $\mathbf{n}_i$ represents the noise present in pixel i, and K is the number of endmembers. We assume Gaussian noise, as do [4, 5]. In general, the number of endmembers K is unknown and chosen arbitrarily depending on the application and the knowledge of the scene. However, an incorrect estimation of this number can dramatically affect the accuracy of the unmixing [5]. Thus, an important first step is to estimate the number of endmembers (or constituents) K that are actually contained in the image, using the observed image rather than any prior knowledge of the scene. The intrinsic dimension is a good first estimation for the number of endmembers.

We consider the eigenvalues of the observation covariance matrix S, defined by

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

where $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$. It is extremely difficult to distinguish between a small signal eigenvalue and a large noise eigenvalue, and several methods have been developed to address this problem. We will consider: Malinowski's Empirical Indicator Function (EIF) [6] [1]; the Harsanyi-Farrand-Chang (HFC) method [3]; and the Random Matrix Theory (RMT) method [2]. However, these methods proved to be sensitive to the structure of the noise in the image [2, 3]. Chang *et al* [3] modified HFC to include a noise-whitening pre-processing step, and we will investigate the effect of including this step in all the methods.

2. METHODS

2.1. Noise Whitening

Roger [7] used a residual based estimation in order to approximate the noise covariance matrix. Chang *et al* use

This research is part of the Centre for High Performance Computing flagship project: Computational Research Initiative in Imaging and Remote Sensing. M. Sears and A. Robin thank CHPC for support. K. Cawse would like to thank S Damelin, K. Wessels, F. van den Bergh, and R. Mathieu for their supervision, as well as the National Research Foundation for funding this research.

this method [6] to estimate the noise covariance $W = diag\{1/\zeta_1^2, \ldots, 1/\zeta_p^2\}$, where $\{\zeta_j^2\}$ are the diagonal elements of S^{-1}. Then the sample covariance may be whitened by:

$$S_W = W^{-1/2} S W^{-1/2}. \quad (3)$$

The resulting noise becomes uncorrelated and its variance equal to 1 in each band [6].

2.2. Noise removal

As described in [8], the noise may be approximated on a per pixel basis, using multiple regression theory as follows:

Let X^i be a column vector containing all pixel values at band i. Let $X^{\partial i}$ be a $(N \times (p-1))$ matrix, where $X^{\partial i} = [X^1, \ldots, X^{i-1}, X^{i+1}, \ldots, X^p]$. Then the pixel values for each band i can be expressed in terms of the pixel values for all other bands, so that $X^i = X^{\partial i}\beta_i + \epsilon_i$, where β_i is the regression vector and ϵ_i is the modeling error. This error, ϵ_i, may be used to approximate the noise, per pixel, in the i^{th} band. (Dias and Nascimento [8] use $\beta_i = [X^{\partial i}]^{\#} X^i$, where $^{\#}$ indicates the pseudo-inverse.)

2.3. Random Matrix Theory

This method assumes i.i.d. noise, and we will attempt to improve the results using the methods described above.

In Random Matrix Theory, research has been done into the largest eigenvalue of a Random Matrix. Since we are assuming Gaussian noise, the largest observed eigenvalue due to noise can be thought of as the largest eigenvalue in a Random Matrix. According to Johnstone [9], the largest eigenvalue of a Random Matrix satisfies the following condition with probability 1 in measure:

$$\lambda \leq \sigma^2(\mu_{N,p} + s(\alpha)\sigma_{N,p}) \quad (4)$$

where σ 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 [4] $\alpha = 0.5\%$). This inequality holds when p and N tend towards infinity, and $p/N = c$ fixed [9]. Then, for real valued data,

$$\mu_{N,p} = N^{-1}(\sqrt{N-0.5} + \sqrt{p-0.5})^2 \quad (5)$$

$$\sigma_{N,p} = N^{-1}(\sqrt{N-0.5} + \sqrt{p-0.5}) \times (\sqrt{N-0.5}^{-1} + \sqrt{p-0.5}^{-1})^{1/3} \quad (6)$$

Note that these parameters do not depend on the number of endmembers, K, and they are fully determined by N and p, which are known.

Then, if the eigenvalues $\{\lambda_j\}_{j=1,\ldots,p}$ of the non-centered observation covariance matrix are sorted in descending order, so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, K is defined as the largest index such that, for all $j \in \mathbb{Z}, 1 \leq j \leq K$, $\lambda_j > \sigma^2(\mu_{N,p} + s(\alpha)\sigma_{N,p})$. This method assumes i.i.d noise, and was shown in [2] to give mostly poor results in application to a real hyperspectral image. However, if we first whiten the data, then the noise variance becomes 1 in each band, fulfilling the i.i.d. criterion. Therefore the sorted, whitened signal eigenvalues may be tested, for all j, by

$$\lambda_j > \mu_{N,p} + s(\alpha)\sigma_{N,p} \quad (7)$$

Note that in the case of hyperspectral imagery, N is very large in proportion to p, so $\sigma_{N,p}$ in equation (6) becomes very small, and therefore the right hand side of equation (4) is not sensitive to the choice of the confidence interval α. For all experiments, α is fixed at 0.5% and so it will not be regarded as a user-determined threshold. (Different values for α were tested, with no effect on the results.)

2.4. Malinowski's EIF

As in RMT, this method assumes i.i.d. noise, and we will evaluate the effect of whitening on this method. Malinowski [1] derived functions to determine IE (Imbedded Error) and IND (Indicator function) in order to evaluate if a matrix is factor analyzable. His application is in the field of chemistry, but Chang and Du [6] apply this to hyperspectral images with some success. Chang and Du define Malinowski's Empirical Indicator Function as:

$$EIF(q) = \frac{(\sum_{j=q+1}^{p} \lambda_q)^{\frac{1}{2}}}{N^{\frac{1}{2}}(p-q)^{\frac{3}{2}}} \quad (8)$$

$$K_{EIF} = \arg\min_q \{EIF(q)\}, \quad (9)$$

where $\{\lambda_q\}$ are the sorted (descending) eigenvalues of the observation covariance matrix. In [6],

$$RE(q) = (\sum_{j=q+1}^{L} \lambda_q)^{\frac{1}{2}} N^{\frac{-1}{2}} (L-q)^{\frac{-1}{2}} \quad (10)$$

$$IND(q) = RE(q)(L-q)^{-2} \quad (11)$$

Malinowski [1] states that the origins of Equations (8) and (9) are unknown. This method assumes i.i.d. noise, and Chang and Du [6] did whiten the data, with improved results, but the authors concluded that the results were not good enough (overestimation) on the real experiments that were tested.

2.5. Harsanyi-Farrand-Chang

HFC does not assume i.i.d. noise, but Chang [6] states that the results may be improved by the whitening procedure discussed above.

This method is based on the premise that the eigenvalues of the centered and non-centered covariance matrices will become the same when the data is centered at zero. Since we assume that the noise has zero mean, this means that the two

sets of eigenvalues will converge at the point where signal ends and noise begins. However, computationally, the difference between the two sets is never zero, and so a false-alarm probability P_F is derived in order to find the threshold.

$$P_F = \int_{\tau}^{\infty} p_0(z)dz \quad (12)$$

$$P_D = \int_{\tau}^{\infty} p_1(z)dz, \quad (13)$$

where τ is the detection threshold. The false alarm probability is fixed (user-determined) to determine τ, and the detection power P_D is maximized. Once τ has been determined, the two sets of eigenvalues are said to converge if they differ by less than this threshold, and their convergence determines K. Note that for each band, the two eigenvalues are compared with a different τ. We calculate τ using the integrals given above, given that $p_0(z) \approx N(0, \sigma^2_{z_L})$, and $\sigma^2_{z_L} \approx \frac{2}{N}(\hat{\lambda}^2_l + \lambda^2_l)$, where the two λ values represent sorted eigenvalues from the centered and non-centered covariance matrices.

More sophisticated noise approximations may be used, and a whitening process is used as a pre-processing step. This method performed the best of all algorithms analysed in Wu *et al*, with its only issue being its dependence on a user-determined threshold.

2.6. Signal Subspace Estimation

This method was introduced by Dias and Nascimento [8], and removes the noise, as opposed to whitening it. They use multiple regression theory to accurately determine the noise value per pixel (described above). The signal covariance matrix is calculated by subtracting the noise vector from the sample vector at each pixel, and using the difference to form a covariance matrix. The intrinsic dimension of the image is then calculated as follows,

$$K_{SSE} = \arg \min_{1 \le i \le p} \{(r^T P'(i) r + 2trace[P(i) C_n / N])\} \quad (14)$$

where r is the mean pixel vector over the image, $P(i) = [e_1, \ldots, e_i]$, where $\{e_j\}$ are the endmembers ordered (descending) by the singular values of the estimated signal correlation matrix, $P'(i) = [e_{i+1}, \ldots, e_p]$, and C_n is the noise covariance matrix.

This method has the advantage of not requiring any user-determined thresholds, and was shown to provide accurate results on synthetic and real data [8].

3. EXPERIMENTS

We have compiled a synthetic dataset, made up by randomly selecting 5 out of 18 minerals chosen from the JPL spectral library. The proportions of each endmember in each pixel are random, with the only restrictions being the non-negative and sum-to-one conditions that are enforced on the proportions. Then, by iterating this method, testing may be done on many different "images", since the endmember combinations and proportions differ for each iteration, and both are randomly selected from a uniform distribution for each iteration. We also varied our values for p, N, and σ in these images.

When creating the synthetic set, with uncorrelated, non-i.i.d. noise, we tested the accuracy of the noise removal method for determining the noise in a particular pixel, and we tested the accuracy of the noise whitening matrix with regards to approximating the noise variance in each band. The removal method seemed a more accurate approximation per pixel (Figure 1 shows the $N(0, \sigma^2)$ distribution in a single pixel) than the whitening method over the image (Figure 2 shows the approximation to σ^2 in variance per band), but both methods accurately approximated the noise.

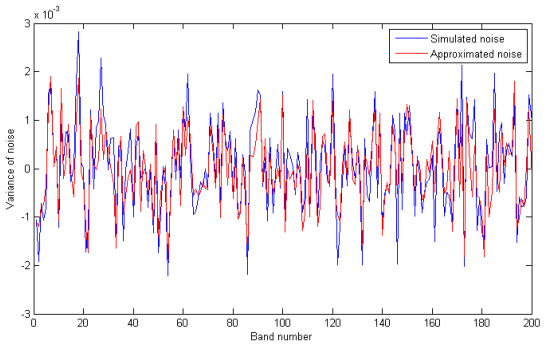

Fig. 1. A comparison between the simulated noise in a pixel (blue), to the approximated noise in a pixel (red) using the noise removal method for approximation. Note the good agreement produced by the method.

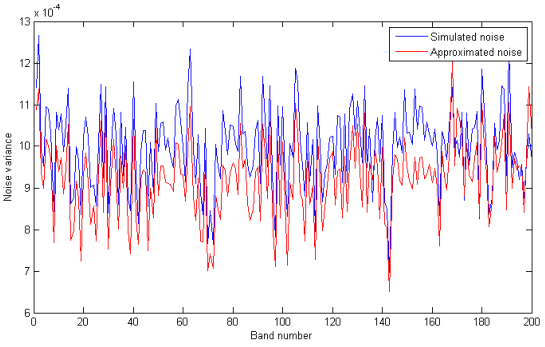

Fig. 2. A comparison between the noise variance per band simulated in an image (blue), to the approximated noise in the image (red) using the noise whitening method for approximation. This also seems to be a good approximation.

Using the noise removal approximation to create a whitening matrix resulted in severe overestimation of K in simulated

Table 1. RMT, EIF, HFC (with $F_D = 10^{-4}, 10^{-3}, 10^{-2}$) and SSE applied to Cuprite, with and without whitening. Where the result is given as *, the method failed to converge. Our whitening method is not applicable to SSE.

	$K_{Original}$	$K_{Whitened}$
RMT	*	37
EIF	75	25
HFC	22, 24, 30	19, 24, 27
SSE	27	-

sets, when using RMT and EIF. When the noise variance was calculated on the image scale, and used in the same way as the whitening method, then accurate K was achieved for both the i.i.d. methods being tested. The noise removal approximation method used in SSE was not applicable to HFC and vice versa. This is because SSE relies on explicit approximations per pixel, and the off diagonal elements in the estimate of the noise covariance matrix using the noise removal approximation seemed to cause errors in HFC, which were minimised by the forced diagonal matrix in the whitening method.

When the whitening method was applied to RMT and EIF, the methods were accurate up to the same noise levels as HFC (standard deviation 0.025 for mean signal values ≈ 0.5), for a test image with 10,000 pixels and 200 bands. The methods were also accurate for variable noise per band, with difference in standard deviation up to 3 times the mean standard deviation, at the worst noise levels. So the simulated tests show positive results for whitening the i.i.d. methods.

We also tested our methods on an AVIRIS flight scene collected over Cuprite, Nevada in 1997. This dataset is available online[1]. The image contains 350×350 spatial pixels, with 189 spectral bands. Wu *et al.* [5] determined the number of endmembers in this scene to be between 22 and 28. Also, ground truth collected by Swayze *et al.* found at least 18 substances [10] (which may not include rarer minerals).

Dias and Nascimento [8] specifically compared SSE and HFC (whitened) under a number of different scenarios, including a smaller subset of Cuprite, and found that SSE is more likely to pick up rare minerals, so its slightly higher value for K is likely to be more accurate. As may be seen in Table 1, the whitened RMT slightly overestimates K, but the results are significantly improved from the failure to converge in the original case. The whitening success can be seen in EIF, which is comparable ($K = 25$) to SSE ($K = 27$) and Chang ($K = 19 - 27$). This method is also computationally less expensive than either method, particularly SSE.

4. CONCLUSION

Methods assuming i.i.d. noise have often been avoided for application to hyperspectral imagery, since this assumption is mostly incorrect for real hyperspectral images [6]. We have shown that a simple whitening pre-processing step eliminates the i.i.d. assumption in two methods, particularly EIF, and these methods are applicable to hyperspectral data, with accurate results in real images. The advantage of these methods is that they are computationally very fast, and simple to implement. They do not rely on user-defined thresholds, and give results comparable with much more complicated algorithms. However, not all whitening methods are applicable, and so this pre-processing step should be chosen with care.

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

5. REFERENCES

[1] R.E. Malinowski, "Determination of the number of factors and experimental error in a data matrix," *Analytical Chemistry*, vol. 49 (4), pp. 612–617, 1977.

[2] K. Cawse, M. Sears, A. Robin *et al.*, "Using random matrix theory to determine the number of endmembers in a hyperspectral image," in *Workshop in hyperspectral image and signal processing: evolution in remote sensing*, 2010.

[3] J.C. Harsanyi, W. Farrand, C-I. Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," *ASPRS*, pp. 236–247, 1994.

[4] S. Kritchman, B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics and Intelligent Laboratory Systems*, vol. 94, pp. 19–32, 2008.

[5] C-C. Wu, W. Liu, C-I. Chang, "Exploration of methods for estimation of number of endmembers in hyperspectral imagery," *Proc. of SPIE*, vol. 7(43), pp. 1–11, 2006.

[6] Chein-I Chang, Qian Du, "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery," *IEEE Transactions of Geoscience and Remote Sensing*, vol. 42 (3), pp. 608–619, 2004.

[7] R.E. Roger, "Principal components transform with simple, automatic noise adjustment," *International Journal of Remote Sensing*, vol. 17 (14), pp. 2719–2727, 1996.

[8] J.M. Bioucas-Dias, J.M.P. Nascimento, "Estimation of signal subspace on hyperspectral data," in *Proceedings of SPIE*, 2005, vol. 5982, pp. 191–198.

[9] I.M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *The Annals of Statistics*, vol. 29(2), pp. 295–327, 2001.

[10] G. Swayze, R. Clark, S. Sutley, A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Summaries of the 3rd Annual JPL Airborne Geosciences Workshop*, 1992, pp. 47–49.