Anomaly Detection and Estimation in Hyperspectral Imaging using Random Matrix Theory tools

Eugénie Terreaux*, Jean-Philippe Ovarlez†, and Frédéric Pascal*

*L2S/CentraleSupélec-CNRS-Université Paris-Sud, 3 rue Joliot-Curie, 91190 Gif-sur-Yvette, France
(e-mail: eugenie.terreaux, frederic.pascal @l2s.centralesupelec.fr)
† ONERA, DEMR/TSI, Chemin de la Hunière, 91120 Palaiseau, France and SONDRA/CentraleSupélec
(e-mail: jean-philippe.ovarlez@onera.fr)

***Abstract*—Anomaly detection aims to detect sources with different spectral characteristics from the background in an hyperspectral image. Classical tools for anomaly detection and estimation are known to have poor performance when they are used on high dimensional hyperspectral image since typically both the number of available sample and their size are large for this kind of imaging. New estimation methods for the number of anomalies, adapted to large dimensional systems, are required. This article points out the limits of classical methods such as Akaike Information Criterion (AIC) or Minimum Description Length (MDL) criteria and it proposes a new estimator based on Random Matrix Theory results better adapted for hyperspectral imaging. Finally, the proposed method is validated on both Monte-Carlo simulations and on experimental data.**

I. Introduction

A hyperspectral image is made of hundreds of images corresponding to the same spatial area but for different wavelengths. This kind of imaging is particularly informative especially since the spectral resolution of these images is important. Thus, it enables to evaluate the kind of material or object present on an image more precisely than with only one image in one spectral band provided the materials spectral characteristics are known [1]. Since hyperspectral images contain a wide range of information, this requires adapted statistical tools. According to the application, different techniques are used to process these information. Among them, unmixing problems [2] and detection issues [1] are often encountered. Unmixing consists in extracting the different materials that are present in one area of the hyperspectral image when the spatial resolution is not sufficiently precise. As for detection issues, which aim at finding a particular object on a particular background, anomalies [3] are to be distinguished from target: an anomaly is characterized by a statistical break in the background, and differs from target detection by the lack of knowledge of the source (the target or the anomaly) spectral characteristics.

This paper focuses on the anomaly detection problem. In hyperspectral imaging, this problem is valuable in multiple fields such as defense (surveillance) [4], environment (rare mineral localisation, etc), astronomy (with spectroscopy), etc. (see e.g. [1], [5]).

The commonly used statistical model for anomaly detection problems is a Gaussian modeling [6] of dimensions (m, N), where N corresponds to the spatial dimension (number of observations) and m is the spectral dimension (size of the observation vector). This Gaussian process drives the statistical background of the hyperspectral image while when unknown anomalies are present, it modifies this background. One way to detect an anomaly is to consider the associated binary hypothesis test and derive the corresponding Likelihood Ratio Test (LRT) for the model. This consists in comparing the Mahalanobis distance of the observation vector to a threshold λ [7]. However, in practice, the properties of the Gaussian noise, namely its statistical mean and its covariance matrix are unknown and estimators are required. Classically, the corresponding Maximum Likelihood Estimators (MLE) are used, i.e. the Sample Mean Vector (SMV) and the Sample Covariance Matrix (SCM). Then the Mahalanobis distribution follows a T^2 Hotelling distribution [8] and characteristics of the test, such as the Probability of False Alarm (PFA) can be estimated. But if m and N are large, with $c_N = m/N$ not small and $c_N \to c$, $c > 0$, Random Matrix Theory (RMT) shows that the Mahalanobis distance is no longer T^2-distributed, making difficult to set the threshold λ for a given PFA. Moreover, the SCM does not converge to the true covariance matrix [9]. In such cases, performance of classical methods like the Minimum Description Length [10], [11] and AIC criteria [12], used to estimate the number of anomalies present in the scene, are degraded.

To fill this gap, in this paper, we propose to use recent results of RMT to derive a better estimator for the number of anomalies. More precisely, we propose an efficient way of setting the corresponding hypothesis test that allows to estimate the number of anomalies. [13] compares other classical methods with a RMT method but without introducing a hypothesis test. The paper is organized as follows: Section II introduces the problem formulation and presents the contribution of this work, i.e. the algorithm for estimating the number of sources in cases of both white and correlated noise. Then, Section III presents some Monte-Carlo simulations and experiments on real data that validate the proposed methodology. Finally, Section IV draws some conclusions and perspectives of this work.

Notations : vectors and matrix are in boldface, matrix in capitals and vectors in small letters, ^{H} the Hermitian operator and ^{T} the transpose.

II. Estimation of the number of anomalies

The main objective for anomaly detection is to decide if there are some anomalies present in a set of N spectral m-vectors (which can belong to the whole hyperspectral image or only to a part of this image) and to estimate their number, that is, the Intrinsic Dimension, defined as the dimension of the signal subspace [13]. In other words, the problem is to detect K anomalies among N observations. Consequently, we consider the following statistical model based on the set of N observation vectors $\mathbf{y}_i$:

- We assume that there are K' observations $\mathbf{y}_i$ containing anomalies, let's say the K' first ones, i.e. for $i = 1, \ldots, K'$, one has,

$$\mathbf{y}_i = \sum_{j=1}^{K_i} \alpha_j\, \mathbf{p}_j + \mathbf{x}_i\,, \qquad i \in [1, K']\,,$$

where each $\mathbf{p}_j$ is the unknown m-vector characterizing the spectral information of the j^{th} anomaly with amplitude α_j and where $\{\mathbf{x}_i\}_{i=1,N}$'s are assumed to be m-dimensional independent complex[1] zero-mean Gaussian vectors with covariance matrix $\boldsymbol{\Sigma}$. The total number K of anomalies leads to the following constraint $\sum_{i=1}^{K'} K_i = K$

- Other observations contain only noise, i.e.

$$\mathbf{y}_i = \mathbf{x}_i \qquad, i \in [K'+1, N]\,.$$

In this paper, we propose two methodologies to estimate K when $\boldsymbol{\Sigma} = \sigma^2\,\mathbf{I}$ and $\boldsymbol{\Sigma} \neq \sigma^2\,\mathbf{I}$ respectively.

Let us consider the SCM of the N-sample $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$, defined as

$$\hat{\mathbf{M}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\,\mathbf{y}_i^H,$$

and its ordered eigenvalues $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{m-1}$. Many techniques based on the eigenvalues of the SCM can be used for estimating the number of sources. The two well known LRT techniques are AIC [12], [14] and MDL [10] but are only valid in the classical asymptotic regime (fixed m and $N \to \infty$).

The proposed approach is to consider, in the large dimensional regime ($m, N \to \infty$ with $m/N \to c > 0$ and for fixed K), the following set of multiple hypothesis test for $k \in \{0, \ldots, \min(m, N) - 1\}$ [15]:

$$\begin{cases} H_0: & \text{at most } k \text{ anomalies present} \\ H_1: & \text{at least } k+1 \text{ anomalies present} \end{cases}$$

Interestingly, the PFA associated to each test which is the probability to detect at least $k+1$ anomalies whereas there are at most k anomalies, does not depend on K. Indeed, the test is based the knowledge of the distribution of the empirical eigenvalues $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{m-1}$ of $\hat{\mathbf{M}}$ when no anomaly is present. So, the PFA is the same for all K.

A. White Gaussian noise

Let us first assume that $\boldsymbol{\Sigma} = \sigma^2\,\mathbf{I}$. In this case, the test consists in comparing the k^{th} eigenvalue $\hat{\lambda}_k$ of the SCM $\hat{\mathbf{M}}$ to a threshold [16]:

$$\hat{\lambda}_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \zeta_N\,. \tag{1}$$

Under H_0, the distribution of $\hat{\lambda}_k$, which is the highest "noise" eigenvalue, has been studied and is known to follow the Tracy-Widom distribution, modulated by two parameters : σ_N and b_N, with a rate of convergence of $O(m^{2/3})$. This result allows to theoretically set the threshold for a given PFA as follows [16]:

$$\zeta_N = \hat{\sigma}^2(k)\,\left(b_N + \frac{\sigma_N}{m^{2/3}}\left(F_{TW}^{-1}(1-\alpha)\right)\right)\,,$$

where F_{TW} stands for the CDF of the Tracy-Widom distribution, α the desired PFA, $b_N = (1+\sqrt{c_N})^2$ and $\sigma_N = (1+\sqrt{c_N})^{4/3}\sqrt{c_N}$ with $c_N = m/N$ and where $\hat{\sigma}^2(k)$ is a consistent estimate of σ^2 [15]

$$\hat{\sigma}^2(k) = \frac{1}{m-k}\sum_{i=k+1}^{m} \hat{\lambda}_i\,.$$

TABLE I. $\hat{K}_{est}$ PERFORMANCES FOR PFA $= 0.01$, 100 SIMULATIONS FOR EACH SNR VALUES, 10 ITERATIONS FOR THE THRESHOLD CALCUL, $K = 4, m = 200, N = 400$.

SNR (dB)	32	28	27	26	24	18
$\hat{K}_{estmean}$	4.00	4.00	3.34	2.12	0.47	0
C	0	0	0.66	1.88	3.53	4.00
Var	0	0	0.27	0.25	0.27	0
$\hat{K}_{MDL}$	0.10	0	0	0	0	0
Var_{MDL}	0.09	0	0	0	0	0
$\hat{K}_{AIC}$	4.00	3.64	1.8	0.27	0	0
Var_{AIC}	0	0.23	0.40	0.20	0	0

The estimated number of anomalies $\hat{K}_{est}$ is then given by

$$\hat{K}_{est} = \underset{k}{\operatorname{argmin}}\left(\hat{\lambda}_k < \zeta_N\right)\,.$$

Notice that the PFA is the same for each test since it depends only on largest noise eigenvalue distribution. This approach, although base on hypothesis tests, allows to provide a consistent (in large dimensional regime) estimator of K.

This method is now compared to the classical MDL and AIC criteria. Each anomaly $k \in [1, K]$ is simply modeled by the spectral information $\mathbf{p}_k = \left(1, e^{2i\,\pi\,k/m}, \ldots, e^{2i\,\pi\,k\,(m-1)/m}\right)^T$ (each $\mathbf{p}_k$ is orthogonal to each others, and hence all the anomalies span a K dimensional subspace). For this simulations, we use $\sigma^2 = 1$, $K = 4$ and we set $\alpha_k = \alpha > 0$ for all k. This implies that all anomalies have the same SNR equal to $10\log_{10}(\alpha^2)$.

For performance analysis, one computes the following quantities, for N_s Monte-Carlo simulations:

- $\hat{K}_{estmean} = \dfrac{1}{N_s}\displaystyle\sum_{n=1}^{N_s} \hat{K}_{est}^{(n)}$, where $\hat{K}_{est}^{(n)}$ is the estimate obtained at the n^{th} simulation, $C = \dfrac{1}{N_s}\displaystyle\sum_{n=0}^{N_s}\left|4 - \hat{K}_{est}^{(n)}\right|$,
- $Var = \dfrac{1}{N_s}\displaystyle\sum_{n=0}^{N_s}\left(\hat{K}_{est}^{(n)} - \hat{K}_{estmean}\right)^2$.

Tables I and II reveal the performance of the proposed method compared with the MDL and AIC criteria. For sufficiently powerful anomalies, $\hat{K}_{est}$ is very close to the true value of K even for small PFA, AIC provides also good results for high SNR but its performance decreases for smaller SNR. Finally, MDL needs a very high SNR to have a non-null estimation.

TABLE II. $\hat{K}_{est}$ PERFORMANCES FOR SNR $= 26$ dB, 100 SIMULATIONS FOR EACH PFA VALUE, 10 ITERATIONS FOR THE THRESHOLD COMPUTATION, $K = 4, m = 200, N = 400$.

PFA	1	0.005	0.001
$\hat{K}_{estmean}$	4.56	3.20	3.08
C	0.55	0.80	0.92
Var	0.25	0.30	0.25
$\hat{K}_{MDL}$	0	0	0
$\hat{K}_{AIC}$	1.7	1.7	1.7

Then, Figure 1 displays the estimation of K, namely $\hat{K}_{estmean}$ versus the PFA for the proposed approach, MDL and AIC estimators

[1] Notice that in hyperspectral image each spectral component of a given pixel vector is real and positive as it represents reflectance or radiance. A global mean vector estimation and a simple Hilbert transform can render them zero-mean complex vector.

for $K = 4$ anomalies, each one with a SNR of 26 dB ($\alpha_k = 1.8$). Of course, AIC and MDL results do not depend on the PFA since it is not based on a hypothesis test. The proposed method really improves the estimation accuracy compared to MDL and/or AIC which both under estimate the number of anomalies. Notice that the region of interest in the one for small PFA (close to zero).

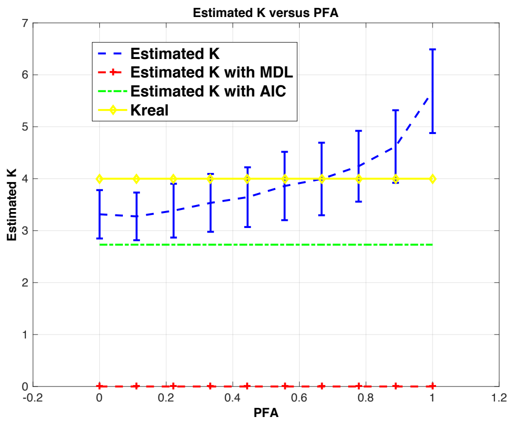

Fig. 1. $\hat{K}_{est}$ versus PFA for $m = 250$, $N = 300$, SNR= 26 dB and $K = 4$ anomalies. In dashed line, the standard deviation for 100 simulations.

B. Correlated Gaussian noise

When the true noise covariance matrix is known, the sample covariance matrix can be whitened by its inverse and the problem fall back to the classical situation of a signal corrupted by a white noise. But in real hyperspectral images, the noise is often correlated and its covariance matrix is not known. Let us now assume that the noise is composed of independent Gaussian vectors, with covariance matrix $\mathbf{\Sigma} \neq \sigma^2 \mathbf{I}$. Let us now consider two cases:

- When a pure noise sequence is available, then it is possible to estimate the covariance matrix with the SCM and whiten the signal. In this case, the model is a so-called F-matrix, and does not provide a Marcenko Pastur limit, so the test given by (1) no longer holds and results of Section II cannot be be theoretically applied. However, it was shown in [17] that, for proper corrected versions of b_N and σ_N, the test (1) was still valid. Another algorithm is developed in [18], based on the decomposition of the covariance matrix.
- No pure noise sequence is available. In this case, one has to propose another approach. The idea is to find a gap between the distances of two consecutive eigenvalues provided that under some assumptions this gap takes place between the highest noise eigenvalue and the lowest anomaly eigenvalue.

Using the assumptions of [9], the strongest is that the signal and the noise cannot be simultaneously correlated, the estimator of K is as follows:

$$\hat{K}_N = \arg \max_{k \in \{0,\dots,L-1\}} \left(\frac{\hat{\lambda}_{k-1}}{\hat{\lambda}_k} > 1 + \varepsilon \right), \text{ with } L \geq K \quad (2)$$

and $\hat{\lambda}_{-1} = +\infty$.

However, the threshold ε can not be theoretically obtained as previously and has to be heuristically estimated. In this work, we propose a way of deriving this threshold for a given PFA. The only PFA of interest is the one associated to the test where $k = K$ with K the true number of anomalies. It is important to notice that this test does not depend on K. Consequently, the PFA can be calculated for $K = 0$ and the threshold is empirically computed from only noise eigenvalues for a given PFA, as explained in the next section.

III. Experimental results

A. Monte-Carlo simulations

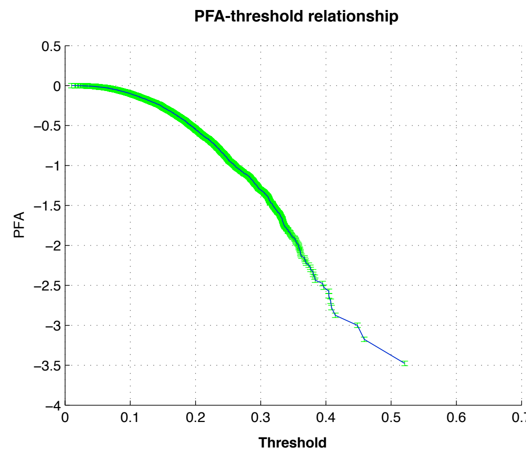

Fig. 2. PFA ($\log_{10}$ scale) versus threshold ε (equation (2))

This section presents some Monte-Carlo simulations to validate the previous results. First, let us consider a correlated Gaussian noise for which a pure-noise sequence is available. The results are contained in Table III. The observation are whitened thanks to the SCM computed on the pure noise sequence.

TABLE III. $K = 4$, PFA = 0.01, m = 400, N = 4000

SNR	44	45	46	46.4	47
$\hat{K}_{estimean}$	0	1.2	3.1	4	4
Var	0	0.16	0.16	0	0
$\hat{K}_{MDL}$	0	0	0	0	0
$\hat{K}_{AIC}$	0	0	2	2.9	4
Var_{AIC}	0	0	0.16	0.16	0

For simplicity reason, we have directly applied the test (non-optimal one) given by (1). First, due to the covariance whitening, one can see that the SNR of each anomaly needs to be higher for achieving similar performance. But again, the proposed test enables to detect the correct number of anomalies from SNR= 46.4 dB while the AIC and MDL underestimate this number. Method proposed for corrected version of b_N and σ_N in [17] should give better results.

Then, when no pure-noise samples are available, Figure 2 displays the PFA versus the threshold ε of equation (2). This is obtained for $K = 0$ (only noise eigenvalues), with 3000 Monte-Carlo simulations of a Gaussian noise, and by counting the number of false detections. Plain lines provide error bars. It is important to notice that the proposed approach allows to set a PFA even if the problem is the estimation of the number of anomalies. This is one of the advantages of the proposed method that allows to provide a confidence criterion with this PFA. In conclusion, although there is no theoretical way of setting the threshold, except using an empirical approach, it enables to detect the correct number of anomalies if they have a sufficiently high SNR. The MDL and AIC methods performance strongly degrades for high values of m and N but also when the noise is strongly correlated.

B. *Application on a real hyperspectral image*

Hyperspectral image under test is plotted on Figure 3 where one can see a car on a road, with around, some vegetation and a ground different from the road.

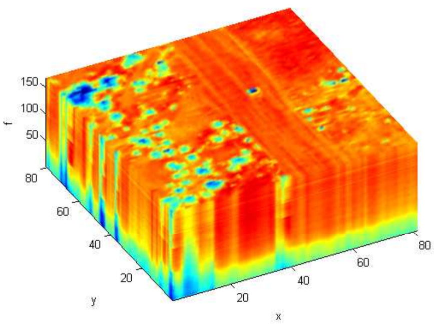

Fig. 3. Hyperspectral image of size $m = 167, N = 81 \times 81$.

In this part, the goal is to test the proposed method considering cars as anomalies.

In order to set an adapted threshold for a selected PFA, the car on the road is removed and replaced with the surrounding pixels. The remaining image is cut into sliding windows of size 21×21 on which covariance matrix and their eigenvalues are estimated. Afterwards the PFA is evaluated for each threshold using all the eigenvalues. Then, the original image (including the car) is tested (equation (2)) using the threshold corresponding to the chosen PFA.

TABLE IV. SUCCESSIVE EIGENVALUES RATIOS FOR THE IMAGE OF SIZE $81 \times 81 \times 167$.

	First eigenvalues ratios					
with car	37	3.4	4.3	3.6	1.1	2.0
without car	16	3.1	4.3	3.3	1.3	2.0

Table IV contains the first successive eigenvalues ratios for the image with and without the car. Notice that other anomalies are present on the image, this explains the important values of these ratios even without the car. However, for a PFA=0.05, the empirical threshold is equal to 34.7 which leads to the detection of one anomaly, namely the car. Finally, AIC leads to the detection of 136 anomalies while MDL detects 126 anomalies. This strong overestimation is due to the size of the data (see [19] for more details) and highlights the interest of having alternative techniques for hyperspectral images.

IV. CONCLUSION

Classical methods for anomaly detection are not adapted for large m and N. In this paper, two hypothesis tests based on recent results of RMT are presented: the first one for a white Gaussian noise and the second for correlated observations. The latter enables to set the required threshold for a given PFA. The theoretical improvement provided by these methods have been illustrated through Monte-Carlo simulations and on a real hyperspectral image. The first results show the interest of the proposed methods compared to the classical MDL and AIC approaches. Further works will address the problem of correlated and non-Gaussian noise.

ACKNOWLEDGMENT

This work has been partially supported by the ICODE institute, research project of the Idex Paris-Saclay.

REFERENCES

[1] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, “Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms,” *Signal Processing Magazine*, 2014.

[2] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2012.

[3] C.-I. Chang and S.-S. Chiang, “Anomaly detection and classification for hyperspectral imagery,” *Geoscience and Remote Sensing, IEEE Transactions on*, 2002.

[4] M. Eismann, A. Stocker, and N. Nasrabadi, “Non-parametric detection of the number of signals: hypothesis testing and random matrix theory,” *Proceedings of the IEEE*, 2009.

[5] ——, “Detection algorithms for hyperspectral imaging applications,” *Signal Processing Magazine*, 2002.

[6] J. Fontera-Pons, M. Veganzones, F. Pascal, and J.-P. Ovarlez, “Hyperspectral anomaly detectors using robust estimators,” *Selected Topics in Applied Earth Observation and Remote Sensing, IEEE Journal of (IEEE-JSTARS), to appear*, 2015.

[7] P. Mahalanobis, “On the generalized distance in statistics,” *Proceedings of the National Institute of Sciences (Calcutta)*, 1936.

[8] J. Fontera-Pons, F. Pascal, and J.-P. Ovarlez, “Adaptative non-zero mean gaussian detection and application to hyperspectral imaging,” *arXiv*, 2014.

[9] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communication*. Cambridge University Press, 2011.

[10] J. Rissanen, “Modeling by shortest data description,” *Automatica*, 1978.

[11] G. Schwartz, “Estimating the dimension of a model,” *Ann. Stat.*, 1978.

[12] M. Wax and T. Kailath, “Detection of signals by information theoric criteria,” *Signal Processing, IEEE Transactions on*, 1985.

[13] A. Robin, K. Cawse-Nicholson, A. Mahmood, and M. Sears, “Estimation of the intrinsic dimension of hyperspectral images: Comparison of current methods,” *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2854–2861, 2015.

[14] H. Akaike, “A new look at the statistical model identification,” *Automatic Control, IEEE Transactions on*, vol. 19, no. 6, pp. 716–723, 1974.

[15] S. Kritchman and B. Nader, “Non-parametric detection of the number of signals: hypothesis testing and random matrix theory,” *Signal Processing, IEEE Transactions on*, 2009.

[16] J. Vinogradova, “Random matrices and applications to detection and estimation in array processing,” Ph.D. dissertation, Telecom ParisTech Paris, 2014.

[17] R. Nadakuditi and J. Silverstein, “Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 3, pp. 468–480, 2010.

[18] K. Cawse-Nicholson, S. Damelin, A. Robin, and M.Sears, “Determining the intrinsic dimension of a hyperspectral image using random matrix theory,” *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1301–1310, 2013.

[19] W. Xu and M. Kaveh, “Analysis of the performance and sensitivity of eigendecomposition-based detectors,” *Signal Processing, IEEE Transactions on*, vol. 43, no. 6, pp. 1413–1426, 1995.