

A CLASS OF ROBUST ESTIMATES FOR DETECTION IN HYPERSPECTRAL IMAGES USING ELLIPTICAL DISTRIBUTIONS BACKGROUND

J. Frontera-Pons¹, M. Mahot¹, J.P. Ovarlez^{1,3}, F. Pascal¹, S.K. Pang², J. Chanussot⁴

ABSTRACT

When dealing with impulsive background echoes, Gaussian model is no longer pertinent. We study in this paper the class of elliptically contoured (EC) distributions. They provide a multivariate location-scatter family of distributions that primarily serve as long tailed alternatives to the multivariate normal model. They are proven to represent a more accurate characterization of HSI data than models based on the multivariate Gaussian assumption. For data in \mathbb{R}^k , robust proposals for the sample covariance estimate are the M -estimators. We have also analyzed the performance of an adaptive non-Gaussian detector built with these improved estimators. Constant False Alarm Rate (CFAR) is pursued to allow the detector independence of nuisance parameters and false alarm regulation.

Index Terms— hyperspectral imaging, target detection, elliptical distributions, M -estimators

1. INTRODUCTION

Anomaly detection and detection of targets or activity such as chemical plumes, aerosols, vehicles, anomalous targets, arise in many military and civilian applications [1]. Hyperspectral imaging sensors provide 2D spatial image data containing spectral information. This information can be used to address such detection tasks. Hyperspectral imaging sensors measure the radiance or reflectance of the materials within each pixel area at a very large number spectral wavelength bands.

It is often assumed that signals, interferences, noises, background are Gaussian stochastic processes. Indeed, this assumption makes sense in many applications. In these contexts, Gaussian models have been widely investigated in the framework of Statistical Estimation and Detection Theory. They have led to appealing and well known algorithms as for instance the Matched Filter and its adaptive variants in radar detection [2, 3]. The mathematical framework for the design and evaluation of detection algorithms is provided by the well known binary hypothesis testing procedure. The basic problem of detecting a complex signal corrupted by an additive noise \mathbf{c} in a m -dimensional complex vector \mathbf{y} can be stated as

a binary hypothesis test with two competing hypotheses H_0 and H_1 . In practice the \mathbf{c} background statistics are unknown and have to be estimated from K secondary data \mathbf{c}_i 's. Under hypothesis H_1 , it is assumed that the observed data \mathbf{y} consists in the sum of a signal $\mathbf{s} = \alpha \mathbf{p}$ and background noise \mathbf{c} , where \mathbf{p} is a perfectly known complex steering vector (characterizing for example the spectral material to detect) and α is the signal amplitude.

Generally, the statistical parameters (covariance matrix \mathbf{M} , mean $\boldsymbol{\mu}$, ...) of the background can be estimated by using all pixels within an area of interest. The size of the area has to be chosen large enough to ensure the invertibility of the covariance matrix and small enough to justify both spectral homogeneity (stationarity) and spatial homogeneity. In hyperspectral imaging, the actual response of a detector to the background pixels differs from the theoretically predicted distribution for Gaussian backgrounds. In fact, as stated in [5], the empirical distribution usually has heavier tails compared to the theoretical distribution, and these tails strongly influence the observed false-alarm rate of the detector. Since the two hypotheses contain unknown parameters (for example, the covariance matrix of the background) that have to be estimated from the data, the detector has to be adaptive, and it is usually designed by using the Generalized-Likelihood-Ratio Test (GLRT) approach.

One of the most general and elegant impulsive noise model often used in radar detection schemes is provided by the so-called *Spherically Invariant Random Vectors* (SIRV). A SIRV \mathbf{y} is a compound process, it is the product of a Gaussian multivariate process $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{M})$ and the square root of a non-negative random scalar variable τ called the *texture*. Thus, the SIRV is fully characterized by the texture (representing an unknown intensity) and the unknown covariance matrix of the Gaussian vector. Another statistical framework is based on the use of Elliptical Random Process which generalizes the SIRV processes. In that framework, the multidimensional vector \mathbf{x} (supposed to be Gaussian in SIRV theory) is here uniformly distributed on the hyper-sphere. That means that a SIRV is also elliptically distributed

One of the major challenging difficulties in SIRV or Spherically distributed random process modeling, is to estimate these two unknown quantities. For example, the classical Sample Covariance Matrix used in adaptive detection in Gaussian noise is not at all the best estimate and does not

¹ : SONDRRA Research Alliance, Supélec, France, ² : DSO National Laboratories, Singapore, ³ : French Aerospace Lab, ONERA DEMR/TSI, France, ⁴ : GIPSA-LAB, Grenoble Institute of Technology, France

correspond to the Maximum Likelihood estimator. In SIRV context, these problems have been investigated in [6] for the texture estimation while [7] and [8] have proposed different estimates for the covariance matrix. A complete statistical analysis of these covariance matrix estimates has been realized in [9]. For Elliptical process, the estimation of the mean and covariance is known as M -estimation theory introduced by [10, 11].

2. ELLIPTICAL DISTRIBUTION

A m -dimensional random complex vector \mathbf{y} is said to have a complex elliptical distribution if its probability density function (PDF) has the form

$$f_{\mathbf{y}}(\mathbf{y}) = |\Sigma|^{-1} h_{\mathbf{y}}((\mathbf{y} - \boldsymbol{\mu})^H \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})), \quad (1)$$

where H denotes the conjugate transpose operator and $h_{\mathbf{y}} : [0, \infty) \rightarrow [0, \infty)$ is any function such as (1) defines a PDF, $\boldsymbol{\mu}$ is the mean vector and Σ is the scatter matrix. The function $h_{\mathbf{y}}$, usually called density generator, is assumed to be only approximately known. Note that it produces density contours corresponding to elliptical surfaces. If the second moments exist, then Σ reflects the structure of the covariance matrix of the elliptically distributed random vector \mathbf{y} , i.e. the covariance matrix equates the scatter matrix up to a scaling constant. We shall denote this complex elliptical distribution by $EC(\boldsymbol{\mu}, \Sigma, h)$. It is worth pointing that the EC class includes an infinity of distributions, notably the Gaussian one, multivariate t distribution or multivariate Cauchy.

3. M- ESTIMATORS

Let $(\mathbf{c}_1, \dots, \mathbf{c}_K)$ be a K -sample of m -dimensional complex independent vectors with $\mathbf{c}_i \sim EC(\boldsymbol{\mu}, \Sigma, h)$, $i = 1, \dots, K$. The complex M -estimators of location and scatter are defined as the joint solutions to the estimating equations:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{\sum_{n=1}^K u_1(t_n) \mathbf{c}_n}{\sum_{n=1}^K u_1(t_n)} \\ \hat{\mathbf{M}} &= \frac{1}{K} \sum_{n=1}^K u_2(t_n^2) (\mathbf{c}_n - \hat{\boldsymbol{\mu}}) (\mathbf{c}_n - \hat{\boldsymbol{\mu}})^H \end{aligned} \quad (2)$$

where $t_n = ((\mathbf{c}_n - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}^{-1} (\mathbf{c}_n - \hat{\boldsymbol{\mu}}))^{1/2}$ and u_1, u_2 are two weighting functions on the quadratic form t_n . Note t_n^2 is in fact, the widely used Mahalanobis distance. M -estimators have first been studied in the real case, defined as solution of (2) with real samples. Existence and uniqueness have been proven in the real case, provided functions u_1, u_2 satisfy a set of general assumptions stated by Maronna [11]. Olilla has shown in [12] that these conditions hold also in the complex

case. M -estimators are particularly suited for estimating the mean vector and the scatter matrix of an elliptical population. When dealing with heavy tailed clutter models, as in HSI, the use of robust estimates decreases the impact of highly impulsive samples and possible outliers.

Remark that if u_1 and u_2 are chosen to be constant and equal to one, the arising estimators correspond to the Sample Mean Vector and Sample Covariance Matrix respectively. They are indeed the the Maximum Likelihood estimators when Gaussian distributions are considered.

We state below two particular estimates belonging to the family of M -estimators. Besides the indicated statistical robustness, they involve some CFAR properties useful for detection issues.

3.1. The Fixed Point estimates

According to the Fixed point approach, the joint estimation of \mathbf{M} and $\boldsymbol{\mu}$ leads to [13]:

$$\hat{\mathbf{M}}_{FP} = \frac{m}{K} \sum_{k=1}^K \frac{(\mathbf{c}_k - \hat{\boldsymbol{\mu}}) (\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H}{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{c}_k - \hat{\boldsymbol{\mu}})} \quad (3)$$

and

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{k=1}^K \frac{\mathbf{c}_k}{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{c}_k - \hat{\boldsymbol{\mu}})}}{\sum_{k=1}^K \frac{1}{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{c}_k - \hat{\boldsymbol{\mu}})}} \quad (4)$$

Obtained when choosing $u_1(s) = s^{-1}$ and $u_2(s) = ms^{-1}$. For the matrix estimate, existence and uniqueness have been established in [14]. Although the proof for simultaneous scatter and location estimates is still an open question, they have been found to be useful and reliable for elliptical contours estimation parameters because of its easy implementation. They are specified by implicit equations and can be easily computed using a recursive algorithm. We refer to [9] for a detailed performance analysis of the Fixed Point covariance matrix estimate. The main results of the statistical properties of the $\hat{\mathbf{M}}_{FP}$ are summarized: $\hat{\mathbf{M}}_{FP}$ is a consistent and unbiased estimate of \mathbf{M} ; its asymptotic distribution is Gaussian and its covariance matrix is fully characterized in [15]; its asymptotic distribution is the same as the asymptotic distribution of a Wishart matrix with $mK/(m+1)$ degrees of freedom.

3.2. The Huber's M -estimates

Using the well-known Huber's ψ function [10] defined as,

$$\psi_k(s) = \min(s, k) \quad (5)$$

with $k > 0$. One can obtain Huber's M -estimator by taking $u_1(s) = \psi_k(s)/s$ and $u_2(s) = \psi_{k^2}/s$. We remark that the Huber function can be seen as a mix between the Fixed Point

estimate and the conventional SCM estimate. Extreme values of t_n^2 outside $[0, k^2]$ are strongly attenuated by the $1/s$ decreasing function (as for the Fixed Point) while normal values below k^2 are uniformly kept (SCM behavior).

4. THE ANMF BUILT WITH THE M-ESTIMATORS

Different types of adaptive non-Gaussian detectors were derived for target enhancement purposes in radar applications. We focus here on the study of the GLRT-Linear Quadratic [16], also known as Adaptive Cosine Estimate,

$$\Lambda(\hat{\mathbf{M}}, \hat{\boldsymbol{\mu}}) = \frac{|\mathbf{p}^H \hat{\mathbf{M}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}})|^2}{(\mathbf{p}^H \hat{\mathbf{M}}^{-1} \mathbf{p})(\mathbf{y} - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad (6)$$

where \mathbf{p} is the spectral steering vector, \mathbf{y} the cell under test and λ the decision threshold. Note that the mean $\hat{\boldsymbol{\mu}}$ is generally omitted in radar detection (and therefore not estimated) as the noise is always zero mean. So, in hyperspectral imaging, as the data represent intensity values and are positive, we need to estimate it, jointly with the covariance matrix \mathbf{M} . Used with the Fixed Point estimates, this detector has essential CFAR properties like texture-CFAR (independent of density generator function), matrix-CFAR (independent of M) and mean-CFAR. Hence, the detector GLRT $\Lambda(\hat{\mathbf{M}}_{FP}, \hat{\boldsymbol{\mu}})$ behaves according to the same distribution regardless of the elliptical distribution used and for different covariance matrices. This is of a major interest in areas of background transition, like for example, in coastal areas (ground and sea) or at the edge of forests (fields and trees) because the detector resulting distribution should be insensitive to the different clutter areas.

4.1. Detector performance

The performance analysis has been realized over the data set provided by DSO National Laboratories, the normalized hypercube is shown in figure 1. The resulting ROC curves (Receiver Operating Characteristic) compare the output of the detector built with the Fixed Point estimates, the Huber M -estimators and the classical SCM. The test conducted consists in placing an artificial target with a fixed SNR through each pixel of the image. For all the possible threshold values, both probability of false alarm and probability of detection are computed. The outcome is illustrated in figure 2.

These preliminary results show the improvement in performance introduced by the use of M -estimators regarding the conventional SCM. The desired robustness properties previously mentioned lead to a higher P_d for small values of the P_{fa} .

4.2. False Alarm Regulation

The ANMF test statistics distribution is well-known for zero-mean Gaussian model. When \mathbf{M} is estimated according to the SCM, it follows a complex Wishart distribution $\mathcal{CW}(K, m; M)$. Taking into account that the Fixed Point

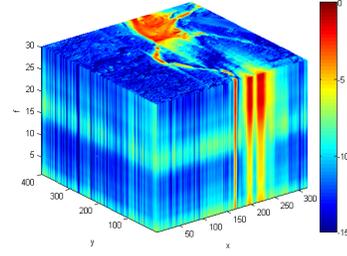


Fig. 1. Normalized data set.

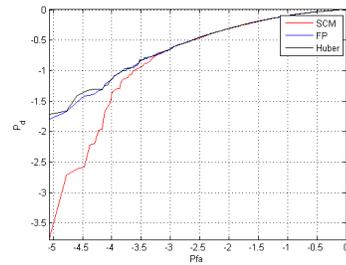


Fig. 2. ROC curves depicting the performance of the detector built with the SCM (in red), the Fixed Point (in blue) and the Huber type (in black) estimates. Probabilities are given in \log_{10} scale.

Matrix asymptotic distribution is the same as the asymptotic distribution of a Wishart matrix with $mK/(m+1)$ degrees of freedom. A theoretical relationship between the detection threshold λ and the Probability of False Alarm $P_{fa} = \mathbb{P}(\Lambda > \lambda | H_0)$ has been stated in [17]:

$$P_{fa} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda) \quad (7)$$

where $a = \frac{m}{m+1}K - m + 2$, $b = \frac{m}{m+1}K + 2$ and ${}_2F_1$ is the Hypergeometric function. This expression holds when $\boldsymbol{\mu}$ is completely known and can be removed from the data. As the joint estimation of \mathbf{M} and $\boldsymbol{\mu}$ is needed, (7) is no longer valid and a gap is evidenced between theoretical and empirical curves.

When $\boldsymbol{\mu}$ is included in the estimation of the SCM, \mathbf{M} is distributed as a complex Wishart with $K-1$ degrees of freedom. The theoretical P_{fa} -threshold relationship for the Gaussian case and SCM estimation has been derived resulting in an expression as in (7) where $a = (K-1) - m + 2$ and $b = (K-1) + 2$. Figure 3 (a) exhibits the regulation of the false alarm for the detector when Gaussian data model is considered.

Simulations held in elliptical distributions context show the early empirical distribution for the detector built with the Fixed Point estimates, hinting its curve analytical expression and its CFAR-texture properties. The figure 4 (b) exemplifies

the empirical results for a K -distribution with shape parameter $\nu = 0.1$. Although a more detailed analysis need to be done.

Note that the previous Pfa-threshold has been derived assuming radar data being complex and is not valid for real data. As the hyperspectral data are real and positive, they have been passed through an Hilbert filter to render them complex.

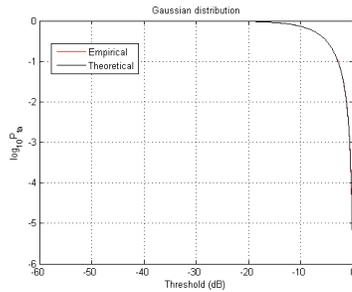


Fig. 3. P_{fa} - threshold relationship for Gaussian model and SCM estimation.

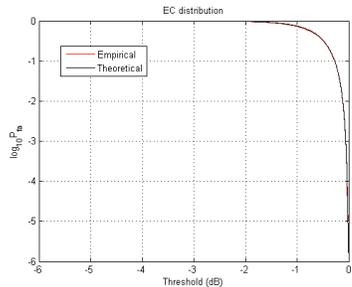


Fig. 4. P_{fa} - threshold relationship for K -distribution with shape parameter $\nu = 0.1$ and FP estimation.

5. CONCLUSIONS

We consider the family of elliptical contoured distributions for impulsive clutter characterization in hyperspectral imaging. In this context, we study different robust estimators for statistical modeling of the background. Particularly, we describe two M estimators, so the Fixed Point and the Huber type, pointing its robust behavior. The main contribution of this article is the analysis performed on the ACE detector when built with these newfangled estimates, showing a better performance in probability of detection terms. Finally we introduce a theoretical relationship for false alarm regulation, when covariance matrix and mean are jointly estimated, foremost for the gaussian case. Empirical simulation results are highlighted for elliptical distribution model.

6. REFERENCES

[1] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *Aerospace and*

Electronic Systems Magazine, IEEE, vol. 25, no. 7, pp. 5–28, 2010.

[2] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans.-AES*, vol. 23, no. 1, pp. 115–127, November 1986.

[3] S. Kraut, L. L. Scharf, and L. T. Mc Whorter, "Adaptive subspace detectors," *IEEE Trans.-SP*, vol. 49, no. 1, pp. 1–16, January 2001.

[4] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 29–43, 2002.

[5] E. Jay, J.-P. Ovarlez, D Declercq, and P Duvaut, "Bord: Bayesian optimum radar detector," *Signal Processing*, vol. 83, no. 6, pp. 1151–1162, June 2003.

[6] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.

[7] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans.-SP*, vol. 50, no. 8, pp. 1908–1915, August 2002.

[8] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans.-SP*, vol. 56, no. 6, pp. 2206–2217, June 2008.

[9] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[10] R. A. Maronna, "Robust M -estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.

[11] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," *Proc. IEEE Workshop on Statistical Signal Processing (SSP), ST Louis, MO*, pp. 445–448, October 2003.

[12] D.E. Tyler, "A distribution-free m -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[13] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum likelihood estimates in compound gaussian noise : Existence and algorithm analysis," *IEEE Trans.-SP*, vol. 56, no. 1, pp. 34–48, January 2008.

[14] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Theoretical analysis of an improved covariance matrix estimator in non-gaussian noise," in *Proc. IEEE-ICASSP*, Philadelphia, March 2005, vol. IV, pp. 69–72.

[15] F. Gini, "Sub-optimum coherent radar detection in a mixture of k -distributed and gaussian clutter," *IEE Proc. Radar, Sonar and Navigation*, vol. 144, no. 1, pp. 39–48, February 1997.

[16] JP Ovarlez, SK Pang, F. Pascal, V. Achard, and TK Ng, "Robust detection using the sirv background modelling for hyperspectral imaging," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 4316–4319.