

Robust ANMF test using Huber's M -estimator

M. Mahot and F. Pascal

SONDRA, Supelec

3 rue Joliot-Curie

91190 Gif-sur-Yvette, France

Email: melanie.mahot@supelec.fr

Email: frederic.pascal@supelec.fr

P. Forster

SATIE, ENS Cachan,

CNRS, UniverSud

61, Av. du Pdt Wilson,

F-94230 Cachan, France

Email: philippe.forster@satie.ens-cachan.fr

J. P. Ovarlez

ONERA, DEMR/TSI

Chemin de la Hunière,

91761 Palaiseau

Cedex, France

Email: jean-philippe.ovarlez@onera.fr

Abstract—In many statistical signal processing applications, the quality of the estimation of parameters of interest plays an important role. We focus in this paper, on the estimation of the covariance matrix. In the classical Gaussian context, the Sample Covariance Matrix (SCM) is the most often used, since it is the Maximum Likelihood estimate. It is easy to manage and has a lot of well-known statistical properties. However it may exhibit poor performance in context of non-Gaussian signals, contaminated or missing data. In that case M -estimators provide a good alternative. In this paper, we extend to the complex data case, a theoretical result already proposed by Tyler in the real data case, deriving the asymptotical distribution of any homogeneous functional of degree 0 of the M -estimates. Then, applying this result to the Adaptive Normalized Matched Filter (ANMF), we obtain a robust ANMF and give the relationship between its Probability of False Alarm (P_{fa}) and the detection threshold.

I. INTRODUCTION

Many signal processing applications deal with the statistical model of collected data. As the latter is traditionally considered to be Gaussian-distributed, one needs only to estimate its mean and covariance matrix. The classical covariance estimator used is the well-known SCM. Indeed it is easy to manage and has a lot of well-known statistical properties. However, the SCM suffers from major drawbacks. Firstly, when the data turn out to be non-Gaussian, as for instance in adaptive Radar and Sonar processing [1], the SCM is a bad estimate: indeed, it performs poorly in the case of impulsive noise and it is not robust to outliers. One possible alternative is given by the so-called M -estimators, originally introduced by Huber [2] and investigated in the seminal work of Maronna [3]. They have been introduced within the framework of elliptical distributions [4] which encompass a large number of well-known distributions as for instance the Gaussian distribution or the multivariate Student (or t) distribution. M -estimators of the covariance matrix are however seldom used in the signal processing community. Notable exceptions are the recent papers by Ollila [5]–[7] who advocates their use in several applications such as array processing.

One possible reason for this lack of interest is that their statistical properties are not well-known in the signal process-

ing community, as opposed to the Wishart distribution of the SCM in the Gaussian context. However, they already have been studied in the real case by Tyler [8]. Thus, to promote the use of M -estimators, we show in this paper that in a Gaussian context and for signal processing estimations problems which only need the covariance matrix up to a scale factor, the estimated parameter of interest has the same mean square error when estimated with the SCM or with an M -estimator with a few more data. Moreover, when the context is non-Gaussian or contains outliers, the performance of the M -estimator is scarcely influenced while the SCM' performance is completely damaged.

In this paper, we extend to the complex data case, the result obtained by Tyler [9] in the real data case. Then we focus on the Adaptive Normalized Match Filter (ANMF) test introduced by Kraut and Scharf [10], [11]. Using the complex analogue of Huber's estimator to estimate the covariance matrix needed in this detector, we obtain a robust ANMF test. Eventually, we give the relationship between the Probability of False Alarm (P_{fa}) and the detection threshold.

This paper is organized as follows. Section II first introduces the required background. Then Section III provides our contribution about the estimators asymptotic distribution. In Section IV, we apply our results to the ANMF test and give the statistical properties of the obtained detector. Eventually, Section V concludes this work.

Vectors (resp. matrices) are denoted by bold-faced lower-case letters (resp. uppercase letters). \sim means "distributed as", $\stackrel{d}{=}$ stands for "shares the same distribution as", \xrightarrow{d} denotes convergence in distribution and \otimes denotes the kronecker product. Moreover, \mathbf{I}_m is the $m \times m$ identity matrix and \mathbf{K} is the commutation matrix which transforms $\text{vec}(\mathbf{A})$ into $\text{vec}(\mathbf{A}^T)$. Eventually, $\text{Im}(\mathbf{y})$ represents the imaginary part of the complex vector \mathbf{y} and $\text{Re}(\mathbf{y})$ its real part.

II. BACKGROUND

In the following section, $'$ represents the transposition operator in the real case, and H in the complex case.

A. Elliptical distribution

Let \mathbf{z} be a m -dimensional real (resp. complex) random vector. \mathbf{z} has a real (resp. complex) elliptical distribution if its probability density function (PDF) can be written as

$$g_{\mathbf{z}}(\mathbf{z}) = |\mathbf{\Lambda}|^{-1} h_{\mathbf{z}}((\mathbf{z} - \boldsymbol{\mu})' \mathbf{\Lambda}^{-1} (\mathbf{z} - \boldsymbol{\mu})), \quad (1)$$

where $h_{\mathbf{z}} : [0, \infty) \rightarrow [0, \infty)$ is any function such that (1) defines a PDF, $\boldsymbol{\mu}$ is the statistical mean and $\mathbf{\Lambda}$ is a scatter matrix. The scatter matrix $\mathbf{\Lambda}$ reflects the structure of the covariance matrix of \mathbf{z} , i.e. the covariance matrix is equal to $\mathbf{\Lambda}$ up to a scale factor. This real (resp. complex) elliptically symmetric distribution will be denoted by $E(\boldsymbol{\mu}, \mathbf{\Lambda})$ (resp. $CE(\boldsymbol{\mu}, \mathbf{\Lambda})$). One can notice that the Gaussian distribution is a particular case of elliptical distributions.

In this paper, we will assume that $\boldsymbol{\mu} = \mathbf{0}_{m,1}$. Without loss of generality *the scatter matrix will be taken to be the covariance matrix*. Indeed, function $h_{\mathbf{z}}$ in (1) can always be defined such that this equality holds.

B. M -estimators of the scatter matrix

Let $(\mathbf{z}_1, \dots, \mathbf{z}_N)$ be a N -sample of m -dimensional real (resp. complex) independent vectors with $\mathbf{z}_i \sim E(\mathbf{0}_{m,1}, \mathbf{\Lambda})$ (resp. $\mathbf{z}_i \sim CE(\mathbf{0}_{m,1}, \mathbf{\Lambda})$), $i = 1, \dots, N$. The real (resp. complex) M -estimator of $\mathbf{\Lambda}$ is defined as the solution of the following equation

$$\widehat{\mathbf{M}} = \frac{1}{N} \sum_{n=1}^N u \left(\mathbf{z}'_n \widehat{\mathbf{M}}^{-1} \mathbf{z}_n \right) \mathbf{z}_n \mathbf{z}'_n. \quad (2)$$

M -estimators have first been studied in the real case, defined as solution of (2) with real samples. Existence and unicity of the solution of (2) has been shown in the real case, provided function u satisfies a set of general assumptions stated by Maronna in [3]. Ollila has shown in [5] that these conditions hold also in the complex case.

Let us now consider the following equation, which is roughly speaking the limit of (2) when N tends to infinity:

$$\mathbf{M} = E \left[u(\mathbf{z}' \mathbf{M}^{-1} \mathbf{z}) \mathbf{z} \mathbf{z}' \right], \quad (3)$$

where $\mathbf{z} \sim E(\mathbf{0}_{m,1}, \mathbf{\Lambda})$ (resp. $CE(\mathbf{0}_{m,1}, \mathbf{\Lambda})$). Maronna has shown in [3] that:

- Equation (3) (resp. (2)) admits a unique solution \mathbf{M} (resp. $\widehat{\mathbf{M}}$) and

$$\mathbf{M} = \sigma \mathbf{\Lambda}, \quad (4)$$

where σ is given in [12].

- A simple iterative procedure provides $\widehat{\mathbf{M}}$.
- $\widehat{\mathbf{M}}$ is a consistent estimate of \mathbf{M} .

C. Wishart distribution

The real (resp. complex) Wishart distribution $W(N, \mathbf{\Lambda})$ (resp. $CW(N, \mathbf{\Lambda})$) is the distribution of $\sum_{n=1}^N \mathbf{z}_n \mathbf{z}'_n$, where \mathbf{z}_n are real (resp. complex circular), i.i.d, Gaussian with zero mean and covariance matrix $\mathbf{\Lambda}$. Let

$$\mathbf{W}_N = N^{-1} \sum_{n=1}^N \mathbf{z}_n \mathbf{z}'_n \quad (5)$$

be the related SCM which will be also referred to, with a slight abuse, as a Wishart matrix. The asymptotic distribution of the Wishart matrix \mathbf{W}_N is

$$\begin{aligned} \sqrt{N} \text{vec}(\mathbf{W}_N - \mathbf{\Lambda}) &\xrightarrow{d} \mathcal{N}(\mathbf{0}_{m^2,1}, (\mathbf{\Lambda} \otimes \mathbf{\Lambda})(\mathbf{I}_{m^2} + \mathbf{K})) \\ \sqrt{N} \text{vec}(\mathbf{W}_N - \mathbf{\Lambda}) &\xrightarrow{d} \mathcal{CN}(\mathbf{0}_{m^2,1}, (\mathbf{\Lambda}^T \otimes \mathbf{\Lambda})) \end{aligned} \quad (6)$$

respectively in the real and complex case.

These few elements being recalled, the main theoretical results are given in the next section.

III. M -ESTIMATORS ASYMPTOTIC PROPERTIES

A. Asymptotic distribution of the complex M -estimators

To derive the asymptotic distribution of the complex M -estimators, several quantities must be introduced. Let $(\mathbf{z}_1, \dots, \mathbf{z}_N)$ be a N -sample of m -dimensional complex independent vectors with $\mathbf{z}_n \sim CE(\mathbf{0}_{m,1}, \mathbf{\Lambda})$, $n = 1, \dots, N$. $\mathbf{z}_n = \mathbf{x}_n + j\mathbf{y}_n$, where $\mathbf{x}_n = \text{Re}(\mathbf{z}_n)$ and $\mathbf{y}_n = \text{Im}(\mathbf{z}_n)$. Moreover $\mathbf{u}_n = (\mathbf{x}_n^T, \mathbf{y}_n^T)^T$.

Let us denote $\widehat{\mathbf{M}}$ the complex M -estimator solution of (2) obtained with the N data \mathbf{z}_n , $n = 1 \dots N$ and the weight function $u(\cdot)$. $\widehat{\mathbf{M}}_u$ is the real M -estimator satisfying equation (2) with the N data \mathbf{u}_n , $n = 1 \dots N$ and the weight function $u_r(s) = u(s/2)$. Moreover, let \mathbf{M}_u be the solution of (3).

Note that $\widehat{\mathbf{M}}_u$ is a $2m \times 2m$ real M -estimator built with independent data which asymptotic distribution has been given by Tyler in [8].

Theorem 3.1: The asymptotic distribution of the complex M -estimator $\widehat{\mathbf{M}}$ is given by

$$\sqrt{N} \text{vec}(\widehat{\mathbf{M}} - \mathbf{M}) \xrightarrow{d} \mathcal{CN}(\mathbf{0}_{m^2,1}, \boldsymbol{\Sigma}), \quad (7)$$

with

$$\boldsymbol{\Sigma} = \sigma_1 \mathbf{M}^T \otimes \mathbf{M} + \sigma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^H, \quad (8)$$

and where σ_1 and σ_2 are the coefficients defined for the asymptotic covariance of the real M -estimator $\widehat{\mathbf{M}}_u$. Their general expressions are given in [8].

Proof: The proof will be given in a forthcoming paper. \blacksquare

In the following part, we extend to the complex case a property obtained by Tyler in [9], for real M -estimators.

B. An important property of complex M -estimators

Let \mathbf{V} be a fixed hermitian positive-definite matrix and $\widehat{\mathbf{V}}$ a sequence of symmetric positive definite random matrix estimates of order m which satisfies

$$\sqrt{N} \left(\text{vec}(\widehat{\mathbf{V}} - \mathbf{V}) \right) \xrightarrow{d} \mathcal{CN}(\mathbf{0}_{m^2,1}, \mathbf{T}), \quad (9)$$

with

$$\mathbf{T} = \nu_1 \mathbf{V}^T \otimes \mathbf{V} + \nu_2 \text{vec}(\mathbf{V}) \text{vec}(\mathbf{V})^H, \quad (10)$$

where ν_1 and ν_2 are any real numbers.

Let $H(\mathbf{V})$ be a r -dimensional multivariate function on the set of $m \times m$ complex hermitian positive-definite matrices, possessing continuous first partial derivatives and such as $H(\mathbf{V}) = H(\alpha \mathbf{V})$ for all $\alpha > 0$. For instance, in DOA methods H is a function which associates an angle to a covariance matrix: $\widehat{\mathbf{V}} \xrightarrow{H} \hat{\theta}$, and in adaptive radar the function which associates the ANMF test statistics to a covariance matrix: $\widehat{\mathbf{V}} \xrightarrow{H} \Lambda(\widehat{\mathbf{V}})$.

Theorem 3.2: The asymptotic distribution of $H(\widehat{\mathbf{V}})$ is given by

$$\begin{aligned} & \sqrt{N} \left(H(\widehat{\mathbf{V}}) - H(\mathbf{V}) \right) \\ & \xrightarrow{d} \mathcal{CN}(\mathbf{0}_{r,1}, \nu_1 H'(\mathbf{V})(\mathbf{V}^T \otimes \mathbf{V}) H'(\mathbf{V})^H). \end{aligned} \quad (11)$$

where $H'(\mathbf{V}) = \left(\frac{dH(\mathbf{V})}{d\text{vec}(\mathbf{V})} \right)$.

Proof: The proof will be given in a forthcoming paper. ■

One can notice that when the data have a complex Gaussian distribution, the SCM is a complex Wishart matrix which verifies the conditions of theorem 3.2. Its coefficients (μ_1, μ_2) are equal to $(1, 0)$. Complex normalized M -estimators also verify the theorem conditions with $(\mu_1, \mu_2) = (\sigma_1, \sigma_2)$. Thus they have the same asymptotic distribution as the complex normalized Wishart matrix, up to a scale factor σ_1 depending on the considered M -estimator.

IV. APPLICATION: THE ROBUST ANMF TEST

In this section we use the complex analogue of Huber's M -estimator as described in [5]. Thus, with the notations of equation (2), the weight function u of this complex M -estimator is such as

$$u(s) = \begin{cases} 1/\beta & \text{for } s \leq k^2 \\ k^2/s\beta & \text{for } s > k^2 \end{cases} \quad (12)$$

where k^2 and β depend on a parameter q , according to

$$q = F_{2m}(2k^2), \quad \beta = F_{2m+2}(2k^2) + k^2 \frac{1-q}{m}. \quad (13)$$

$F_m(\cdot)$ is the cumulative distribution function of a χ^2 variate with m degrees of freedom. Briefly, $q = 1$ leads to the SCM

while smaller values bring robustness to outliers. The chosen value is $q = 0.75$. The theoretical asymptotic covariance of Huber's M -estimator is given by (7).

A. Asymptotic performance of the ANMF test using Huber's M -estimator

We consider an adaptive radar receiving a data vector \mathbf{y} of length m . The estimated covariance matrix of the environment is $\widehat{\mathbf{M}}$ and we try to detect signals with steering vector \mathbf{p} . This steering vector defines the DOA and/or the target speed. The ANMF test statistics is

$$\Lambda(\widehat{\mathbf{M}}|\mathbf{y}) = \frac{|\mathbf{p}^H \widehat{\mathbf{M}}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \widehat{\mathbf{M}}^{-1} \mathbf{p})(\mathbf{y}^H \widehat{\mathbf{M}}^{-1} \mathbf{y})}. \quad (14)$$

In radar detection, to be able to use a detector, one must be able to determine the detection threshold λ which ensures a fixed false alarm rate. The probability of false alarm P_{fa} is defined as the probability of detecting a signal in the noise only case:

$$P_{fa} = \mathbb{P}(\Lambda > \lambda | H_0) \quad (15)$$

where H_0 is the noise only assumption.

The ANMF test statistics distribution is well-known in the Gaussian noise case where $\widehat{\mathbf{M}}$ is the Wishart distributed SCM. As already mentioned, the SCM is a non robust estimator. In this context, the P_{fa} value is given by (see for instance [13]):

$$P_{fa} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (16)$$

where

$$a = N - m + 2, \quad b = N + 2 \quad (17)$$

and ${}_2F_1$ is the Hypergeometric function defined as

$${}_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)x^k}{\Gamma(c+k)k!} \quad (18)$$

As we mentioned previously, robustness may be brought by using M -estimators. The question raised is: how do we correctly set the P_{fa} value? As we will now see, the answer is provided by theorem 3.2. Let us notice first that the test statistics (14) is unchanged if we substitute $\widehat{\mathbf{M}}$ with $\widehat{\mathbf{M}}/\text{Tr}(\widehat{\mathbf{M}})$. Moreover $\widehat{\mathbf{M}}/\text{Tr}(\widehat{\mathbf{M}})$ satisfies the theorem requirements. Therefore, for N large and any elliptically distributed noise, the PFA is still given by (16) if we replace N by N/σ_1 .

B. Simulations using the robust ANMF

In both figures 1 and 2, we have computed $\Lambda(\widehat{\mathbf{M}}|\mathbf{y})$. The vertical scale represents the logarithm of the variance of Λ obtained with the SCM and the complex analogue of Huber's M -estimator defined in (12). The horizontal scale represents the number of samples used to estimate the covariance matrix. In figure 1, we consider a Gaussian context and a third curve

represents the variance of Λ for $\sigma_1 N$ data. As one can see, it overlaps the SCM's curve, illustrating theorem 3.2. Here, σ_1 is equal to 1.066. In figure 2, we have considered a K-distributed environment, with shape parameter firstly equal to 0.1, and then 0.01 for a more impulsive noise. This illustration brings once again to our minds that the SCM isn't robust in a non-Gaussian context contrary to Huber's M -estimator. Indeed, the more the noise differs from a Gaussian noise, the more the detector's variance is deteriorated in the SCM case, while it still gives good results with Huber's M -estimator.

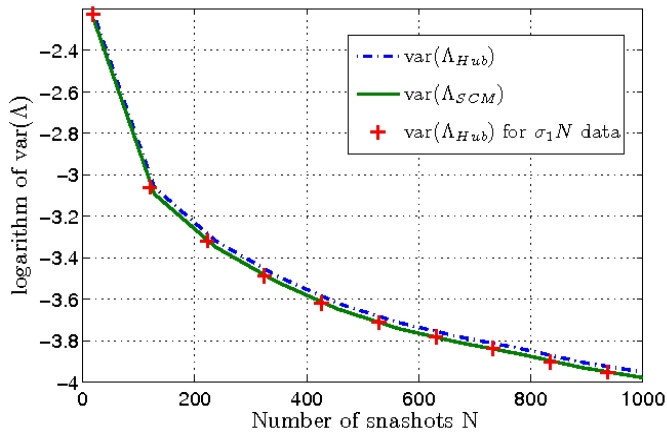


Fig. 1. Variance (logarithmic scale) on the ANMF test statistics for Huber's estimate and the SCM, with a spatially white Gaussian additive noise.

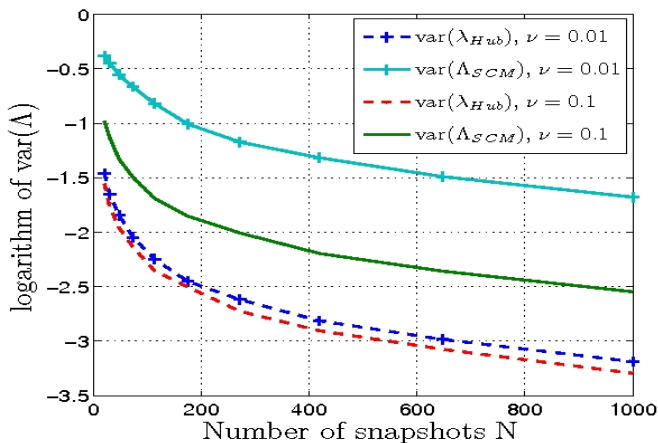


Fig. 2. Variance (logarithmic scale) on the ANMF test statistics for Huber's estimate and the SCM, with a K-distributed additive noise.

V. CONCLUSION

In this paper we have analyzed the statistical properties of general M -estimators of scatter matrix. Using the already known result of the asymptotic covariance of real M -estimators, we have extended these results to the complex

case. Then we have shown that for a Gaussian distribution, parameters which are function of the covariance matrix share the same asymptotic distribution, up to a scalar factor σ_1 , whether they are obtained with a complex M -estimator or the classical SCM. The function H which gives the parameter must be such as $H(\alpha M) = H(M)$. Roughly speaking, the covariance matrix must be needed only up to a scale factor, which is often the case in signal processing applications. Thus, in the Gaussian case M -estimators built with $\sigma_1 N$ data achieve the same asymptotic performance as the SCM built with N data. And in the non-Gaussian case, they are more robust than the SCM. σ_1 is in most cases very close to 1. To illustrate our results, we have simulated a radar detector, the ANMF test and we were able to deduce the PFA-threshold. As a conclusion, this theoretical analysis is in favor of M -estimators since they provide robustness and behave almost like standard tools.

In a forthcoming paper, we will give more information on the scale factor σ_1 , showing that it is most often very close to 1.

ACKNOWLEDGMENT

The authors would like to thank the Direction Générale de l'Armement (DGA) for funding this project.

REFERENCES

- [1] S. M. Kay, *Fundamentals of Statistical Signal Processing - Detection Theory*. Prentice-Hall PTR, 1998, vol. 2.
- [2] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [3] R. A. Maronna, "Robust M -estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.
- [4] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 32, no. 4, pp. 419–430, 1970.
- [5] E. Ollila and V. Koivunen, "Robust antenna array processing using M -estimators of pseudo-covariance," in *Proc. 14th IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, 2003, pp. 7–10.
- [6] —, "Influence functions for array covariance matrix estimators," *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, ST Louis, MO, pp. 445–448, October 2003.
- [7] E. Ollila, L. Quattropani, and V. Koivunen, "Robust space-time scatter matrix estimator for broadband antenna arrays," *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 1, pp. 55 – 59 Vol.1, oct. 2003.
- [8] D. Tyler, "Radial estimates and the test for sphericity," *Biometrika*, vol. 69, no. 2, p. 429, 1982.
- [9] —, "Robustness and efficiency properties of scatter matrices," *Biometrika*, vol. 70, no. 2, p. 411, 1983.
- [10] 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.
- [11] S. Kraut and L. Scharf, "The cfar adaptive subspace detector is a scale-invariant glrt," *Signal Processing, IEEE Transactions on*, vol. 47, no. 9, pp. 2538–2541, 1999.
- [12] M. Bilodeau and D. Brenner, *Theory of multivariate statistics*. Springer Verlag, 1999.
- [13] F. Pascal, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Constant false alarm rate detection in spherically invariant random processes," in *Proc. of the European Signal Processing Conf.*, Vienna, September 2004, pp. 2143–2146.