

# On the Use of Empirical Likelihood for Non-Gaussian Clutter Covariance Matrix Estimation

Hugo Harari-Kermadec<sup>1</sup> and Frédéric Pascal<sup>2</sup>

<sup>1</sup> CREST-LS and University Paris-Dauphine, France,

<sup>2</sup> SATIE, ENS Cachan, CNRS, UniverSud, 61, av President Wilson, F-94230 Cachan, France

email: harari@ensae.fr, pascal@satie.ens-cachan.fr

**Abstract**—This paper presents a improved estimation scheme when the clutter distribution is unknown. The Empirical Likelihood (EL) is a recent semi-parametric estimation method [11] which allows to estimate unknown parameters by using information contained in the observed data such as constraints on the parameter of interest as well as an a priori structure. The aim of this paper is twofold. First, the empirical likelihood is briefly introduced and then, some constraints on the unknown parameters are added. To illustrate this situation, we focus on the problem of estimating the clutter covariance matrix when this matrix is assumed to be Toeplitz [4], [7]. Finally, theoretical results are emphasized by several simulations corresponding to real situations: the mixture of a Gaussian (thermal noise) and a non-Gaussian (clutter) noise.

**Index Terms**—Empirical Likelihood, covariance matrix estimation, non-Gaussian noise.

## I. INTRODUCTION

Non-Gaussian noise characterization has gained many interests since experimental radar clutter measurements, made by organizations like MIT [1], showed that these data are correctly described by non-Gaussian statistical models. One of the most tractable and elegant non-Gaussian model comes from the so-called *Spherically Invariant Random Vector* (SIRV) theory.

A SIRV is the product of a Gaussian random vector with the square root of a non-negative random variable (*texture*). This model leads to many results in terms of parameters estimation as well as in target detection [3], [6], [13].

However, for radar applications, one has to model the noise as the sum of the clutter (due to the ground echoes) and the thermal noise (electronics). But the resulting noise can apart from a SIRV and previous results become suboptimal.

In these cases, since the distribution of the resulting noise is not known, we propose in this paper to estimate the noise covariance matrix, essential for the detection, without priors on the Probability Density Function (PDF) but only by using priors on the parameter (structure, other moments, ...).

The paper is organized as follows. Section II presents the problem formulation while in section III, the Empirical Likelihood (EL) method is briefly introduced. Section IV

provides the main contribution of this paper by establishing several covariance matrix estimates according to priors. Finally, Section V illustrates theoretical results by simulations, corresponding to real situations.

## II. PROBLEM FORMULATION

In this section, we introduce the statistical framework and the notations used in this paper.

### A. Notations

In the following,  $H$  denotes the conjugate transpose operator,  $\top$  denotes the transpose operator,  $E[\cdot]$  stands for the expectation of a random variable,  $E_P[\cdot]$  is the expectation under the data probability  $P$  and  $\text{Tr}(\mathbf{M})$  is the trace of matrix  $\mathbf{M}$ .  $\mathbb{C}$  (respectively  $\mathbb{R}$ ) denotes the set of complex (resp. real) numbers, while for any integer  $p$ ,  $\mathbb{C}^p$  (resp.  $\mathbb{R}^p$ ) represents the set of  $p$ -vectors with complex (resp. real) elements. For  $z \in \mathbb{C}$ , we write  $\text{Re}(z)$  and  $\text{Im}(z)$  its real and imaginary parts.

### B. Statistical Framework

The basic problem of detecting a complex known signal  $\mathbf{s}$  corrupted by an additive noise  $\mathbf{c}$  in a  $p$ -dimensional complex vector  $\mathbf{y}$  can be stated as the following binary hypothesis test:

$$\begin{cases} H_0 : \mathbf{y} = \mathbf{c} & \mathbf{y}_k = \mathbf{c}_k \quad k = 1, \dots, K, \\ H_1 : \mathbf{y} = \mathbf{s} + \mathbf{c} & \mathbf{y}_k = \mathbf{c}_k \quad k = 1, \dots, K, \end{cases} \quad (1)$$

where the  $\mathbf{c}_k$ 's are  $K$  signal-free independent measurements, traditionally called the secondary data, used to estimate the clutter covariance matrix.

This problem conserves all its generality since the  $\mathbf{c}_k$ 's are random variables with unknown PDF:

$$\mathbf{c}_k = \tau \mathbf{M}^{1/2} \varepsilon_k, \quad (2)$$

where  $\tau$  is an unknown (deterministic) power parameter considered as a nuisance parameter and where  $\varepsilon_k$  is zero-mean random  $p$ -vector with identity covariance matrix.

For identifiability considerations, the covariance matrix  $E[\mathbf{c}\mathbf{c}^H] = \tau^2 \mathbf{M}$  has to be normalized. According to [8], we set  $\text{Tr}(\mathbf{M}) = p$ . We will denote  $d$  for the unknown parameter dimension and  $\mathbf{N} = \tau^2 \mathbf{M}$  in the sequel.

The EL method is not based on the true PDF of  $\varepsilon_k$  and thus, can be used in any noise context. For example, one can consider the realistic problem of a mixture of Gaussian and non-Gaussian noises. In the simulations, the secondary data  $\mathbf{c}_k$  will be the sum of a SIRV  $\mathbf{z}_k$  with identity covariance matrix and an independent zero-mean Gaussian  $p$ -vector  $\mathbf{x}_k$  with identity covariance matrix:

$$\mathbf{c}_k = \tau \mathbf{M}^{1/2} \left( \frac{\alpha \mathbf{z}_k + (1 - \alpha) \mathbf{x}_k}{\sqrt{\alpha^2 + (1 - \alpha)^2}} \right), \quad (3)$$

where  $\alpha \in [0, 1]$  is the mixture coefficient that measures the distance to the Gaussian.

### III. EMPIRICAL LIKELIHOOD

The Empirical Likelihood method is a semi-parametric method designed for estimation problems in which the parameter of interest  $\mathbf{N}$  is defined as the solution of an estimating equation.

The covariance matrix estimation problem, that we consider in this paper, can be formulated in such a way:

$$E_{P_0}[\mathbf{N} - \mathbf{c}\mathbf{c}^H] = \mathbf{0}. \quad (4)$$

where  $\mathbf{0}$  denotes the null vector with appropriate dimension ( $n = p^2$  here). Let us recall that  $p$  is the dimension of the observations  $\mathbf{c}$  and  $d$  the dimension of the parameter of interest. Notice that the expectation is taken under the true PDF  $P_0$  of  $\mathbf{c}$  which is assumed to be unknown.

In this paper, we only use the EL method to estimate the clutter covariance matrix. Therefore, we will restrict ourselves to build the corresponding likelihood and to derive its Maximum Empirical Likelihood (MEL) estimate. Nevertheless, it is interesting to notice that the EL method can be extended to the detection problems by using Empirical Likelihood Ratios and confidence areas (see e.g. [11]).

#### A. Likelihood

In this section, we introduce EL methodology from the classical likelihood context. For that pedagogic purpose, we consider the family of multinomials  $G$  charging the data set as if it was a parametric model for the data. Notice that this parametric model assumption will never be made in our method, we just made it to interpret EL as a classical likelihood. All the details and technical arguments are given in Owen's book [12].

We define, for  $G$  and  $\mathbf{N}$  verifying the moment condition  $E_G[\mathbf{N} - \mathbf{c}\mathbf{c}^H] = \mathbf{0}$ , the PDF

$$G(\mathbf{c}) = \begin{cases} q_k & \text{if } \exists k, \mathbf{c} = \mathbf{c}_k, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The corresponding likelihood is called Empirical Likelihood:

$$EL(\mathbf{N}) = EL(\mathbf{c}_1, \dots, \mathbf{c}_K, \mathbf{N}) \quad (6)$$

$$= \sup_{(q_k)} \left\{ \prod_{k=1}^K q_k \left| \sum_{k=1}^K q_k (\mathbf{N} - \mathbf{c}_k \mathbf{c}_k^H) = \mathbf{0}, \sum_{k=1}^K q_k = 1 \right. \right\}.$$

The main technical difficulty, evaluating the empirical likelihood  $EL(\mathbf{N})$  at any given  $\mathbf{N}$ , is resolved by a Lagrangian method and  $EL(\mathbf{N})$  can be written in an easier form:

$$-2 \log(EL(\mathbf{N})) = \inf_{\lambda} \left\{ 2 \sum_{k=1}^K \log \left( K \left( 1 + \lambda^\top (\mathbf{N} - \mathbf{c}_k \mathbf{c}_k^H) \right) \right) \right\}, \quad (7)$$

because the optimal weights write

$$q_k^* = \frac{1}{K} \left( 1 + \lambda^{*\top} (\mathbf{N} - \mathbf{c}_k \mathbf{c}_k^H) \right)^{-1}, \quad (8)$$

where  $\lambda^*$  is the optimal Lagrange multiplier and depends on  $\mathbf{N}$ .

#### B. Maximum Empirical Likelihood without prior

Now,  $\hat{\mathbf{N}}_{EL}$  is define as the arginf of  $EL(\mathbf{N})$ :

$$\hat{\mathbf{N}}_{EL} = \arg \inf_{\mathbf{N}} \left( \inf_{\lambda} \left\{ 2 \sum_{k=1}^K \log \left( K \left( 1 + \lambda^\top (\mathbf{N} - \mathbf{c}_k \mathbf{c}_k^H) \right) \right) \right\} \right). \quad (9)$$

If no restriction is assumed on the structure of  $\mathbf{N}$  (except the Hermitian structure which involves  $d_1 = p^2$  unknown real elements), the Maximal Empirical Likelihood (MEL) estimate is given by the following expression

$$\hat{\mathbf{N}}_{EL1} = \overline{\mathbf{c}\mathbf{c}^H} = \frac{1}{K} \sum_{k=1}^K \mathbf{c}_k \mathbf{c}_k^H \quad (10)$$

where the notation  $\bar{\mathbf{a}}$  is used for the empirical mean of the vector  $\mathbf{a}$ :  $\frac{1}{K} \sum_{k=1}^K \mathbf{a}_k$ .

Notice that  $\hat{\mathbf{N}}_{EL1}$  is equal to the Maximum Likelihood (ML) estimate under Gaussian assumptions.

#### C. Additional priors

This explicit expression (10) for the estimate can be obtain even when some restrictions are made on the covariance matrix structure. The particular case of Toeplitz matrix will be considered in section IV. The estimation scheme can take into account some priors information by introducing an additional equation to Eq. (4):

$$E_{P_0} \left[ \begin{matrix} \text{vec}(\mathbf{N} - \mathbf{c}\mathbf{c}^H) \\ \mathbf{c}^{(prior)} \end{matrix} \right] = \begin{pmatrix} \mathbf{0}_{p^2} \\ \mathbf{0}_s \end{pmatrix}. \quad (11)$$

where  $\mathbf{c}^{(prior)}$  is a vector resulting in a data transformation which reflects the priors information and where the operator  $\text{vec}(\cdot)$  reshapes a  $m \times n$  matrix elements into a  $mn$  column

vector.

By rewriting Eq. (11) as

$$E_{P_0} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{pmatrix} \text{vec}(\mathbf{N}) \\ \mathbf{0} \end{pmatrix}, \quad (12)$$

where  $\mathbf{v} = \text{vec}(\mathbf{c}\mathbf{c}^H)$ ,  $\mathbf{w} = \mathbf{c}^{(prior)}$  and by setting, for the covariance matrix of the vector  $\begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix}$

$$\text{Var} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad (13)$$

one obtains the following closed-form expression for the MEL estimate:

$$\hat{\boldsymbol{\theta}}_{MEL} = \bar{\mathbf{v}} - V_{12}V_{22}^{-1}\bar{\mathbf{w}}. \quad (14)$$

This closed-form can only be obtained when priors information can be written in terms of Eq. (11). This result has been proved in [12], page 52.

Notice that when there is no prior information, i.e.  $\mathbf{w} = \mathbf{0}$ , one obtains  $\hat{\mathbf{N}}_{EL1}$ . This is also the case when  $\mathbf{w}$  is uncorrelated with  $\mathbf{v}$  because  $V_{12}$  is the null matrix.

#### IV. APPLICATION TO TOEPLITZ MATRIX ESTIMATION

In this section, we use priors information on the covariance matrix structure.

Several problems in Signal Processing assume that the covariance matrix of the additive noise has a Toeplitz structure [10], [14], [4], [7]. The covariance matrix (or correlation matrix since  $E_{P_0}[\mathbf{x}] = 0$ ) is often a Toeplitz matrix, since the data vectors consist of subsequent samples from a single signal or times series. The Toeplitz matrix are also met in case of stationary random processes. For instance, because of the stationarity of the input process, the covariance matrix of the autoregressive (AR) process is a Toeplitz matrix. [9] gives a good tutorial on Toeplitz matrices, which contains most of their properties.

Notice that there already exists methods for structured covariance matrix estimation in which the Toeplitz case is treated, see e.g. [2], [5], but these methods are based on a parametric model, usually the Gaussian one.

##### A. First prior: $\mathbf{N}$ is real and has a Toeplitz structure

A first step is to assume that  $\mathbf{N}$  has real valued elements. The dimension of the parameter is therefore reduced from  $p^2$  to  $p(p+1)/2$ .

Now, let us assume that  $\mathbf{M}$  is a Toeplitz matrix: for  $i, j \in \{1, \dots, p\}$

$$\exists(a_0, \dots, a_{p-1}) \in \mathbb{R}^p, \text{ such that } M_{ij} = a_{|i-j|}. \quad (15)$$

In this subsection, one takes advantage of the Toeplitz structure on  $\mathbf{M}$ , and as a consequence on  $\mathbf{N} = \tau^2 \mathbf{M}$ . Then, the constraint integrated in our process is the following: elements of the main diagonal  $N_{ii}$  are all equal to  $\tau^2$  and elements on the upper diagonals  $N_{ij}$ , for  $i < j$ , are equal to  $\tau^2 a_i$ .

This changes both the dimension of the problem and the moment condition. Indeed, the number of unknown real parameters is 1 for each of the  $p$  upper diagonals. Therefore,  $d_2 = p$ . The estimating equation is modified to take the structure into account. For that purpose, we define the function  $m_2$  of dimension  $n_2$  as

$$\mathbf{m}_2(\mathbf{c}, \mathbf{N}) = \begin{pmatrix} (c_i c_{i+j} - \tau^2 a_j)_{1 \leq i \leq p, 1 \leq j \leq p-i} \\ (c_i^2 - \tau^2)_{1 \leq i \leq p} \end{pmatrix}, \quad \text{with } d_2 = p \text{ and } n_2 = \frac{p(p+1)}{2}. \quad (16)$$

This leads to a new estimate  $\hat{\mathbf{N}}_{EL2}$  of  $\mathbf{N}$ , which integrates the constraint on the Toeplitz structure of the real covariance matrix  $\mathbf{M}$ . This writes

$$\hat{\mathbf{N}}_{EL2} = \arg \inf_{(\mathbf{N}, \lambda)} \left\{ \sum_{k=1}^K \log(1 + \lambda^\top \mathbf{m}_2(\mathbf{c}_k, \mathbf{N})) \right\}. \quad (17)$$

We can rewrite the constraint in terms of expectations in order to obtain an explicit form of the estimate by means of Eq. (14). For simplicity purpose, we give the constraints for  $p = 3$ . Let

$$\begin{aligned} \mathbf{v} &= \mathcal{R}e(c_1 c_1^H, c_1 c_2, c_1 c_3)^\top \\ \mathbf{w} &= (\mathcal{I}m(c_1 c_2, c_1 c_3), c_1 c_1^H - c_2 c_2^H, \\ &\quad c_1 c_1^H - c_3 c_3^H, c_1 c_2 - c_2 c_3)^\top \end{aligned} \quad (18)$$

Eq. (14) gives estimates for the first line of  $\mathbf{N}$ :

$$\begin{pmatrix} \widehat{\tau^2}, \widehat{\tau^2 a_1}, \widehat{\tau^2 a_2} \end{pmatrix} = \bar{\mathbf{v}} - V_{12}V_{22}^{-1}\bar{\mathbf{w}} \quad (19)$$

The estimate  $\hat{\mathbf{N}}_{EL2}$  writes then

$$\hat{\mathbf{N}}_{EL2} = \begin{bmatrix} \widehat{\tau^2} & \widehat{\tau^2 a_1} & \widehat{\tau^2 a_2} \\ \widehat{\tau^2 a_1} & \widehat{\tau^2} & \widehat{\tau^2 a_1} \\ \widehat{\tau^2 a_2} & \widehat{\tau^2 a_1} & \widehat{\tau^2} \end{bmatrix} \quad (20)$$

##### B. Second constraint: $\mathbf{N}$ shares a special Toeplitz structure correlation matrix

For some applications, a second assumption can be added. The correlation information contained in the covariance matrix is assumed only through one parameter, the correlation coefficient  $\rho$ :

$$M_{ij} = \rho^{|i-j|} \text{ and } N_{ij} = \tau^2 \rho^{|i-j|}, \quad (21)$$

for  $1 \leq i, j \leq m$  and for  $0 < \rho < 1$ . Notice that the covariance matrix  $\mathbf{M}$  is fully defined by the parameter  $\rho$ , which characterizes the correlation of the data.

Then, the unknown parameters of  $\mathbf{N}$  are 2 real scalar,  $\rho$  and  $\tau$ . Thus, the dimension of the problem is  $d_3 = 2$ , while

the dimension of the estimating equation remains unchanged,  $n_3 = \frac{p(p+1)}{2}$ :

$$\mathbf{m}_3(\mathbf{c}, \mathbf{N}) = \left( (c_i c_{i+j} - \tau^2 \rho^j)_{1 \leq i \leq p, 0 \leq j \leq p-i} \right),$$

with  $d_3 = 2$  and  $n_3 = \frac{p(p+1)}{2}$ . (22)

This leads to the last estimate  $\hat{\mathbf{N}}_{EL3}$  of  $\mathbf{N}$ , defined by

$$\hat{\mathbf{N}}_{EL3} = \arg \inf_{(\hat{\mathbf{N}}, \lambda)} \left\{ \sum_{k=1}^K \log(1 + \lambda^\top \mathbf{m}_3(\mathbf{c}_k, \hat{\mathbf{N}})) \right\}. \quad (23)$$

#### Remark IV.1

Notice that no analytical expression for this last estimate is available, because we have not been able to write the constraints  $c_i c_{i+j} - \tau^2 \rho^j$  as expectations in order to use Eq. (14). Nevertheless, one can give a general expression for all these estimates:

$$\hat{\mathbf{N}}_{ELj} = \sum_{k=1}^K q_k^*(j) \mathbf{c}_k \mathbf{c}_k^H,$$

where the  $q_k^*(j)$  depend on the constraints. For example, in the case of no constraint (i.e.  $\hat{\mathbf{N}}_{EL1}$ ), the  $q_k^*(1)$  are all equal to  $K^{-1}$ . This corresponds to the ML estimate of the covariance matrix for a Gaussian vector.

For the other estimates, the  $q_k^*(j)$  allow to give a weight on the  $k^{th}$  data  $\mathbf{c}_k$  in order to fulfill the a priori conditions. As a consequence,  $\hat{\mathbf{N}}_{EL2}$  has a Toeplitz structure while  $\hat{\mathbf{N}}_{EL3}$  satisfies the special Toeplitz structure given by  $\rho$ .

These theoretical estimates of  $\mathbf{N}$  will be compared in the section V thanks to simulations on their Mean Square Error (MSE).

## V. SIMULATIONS

In order to enlighten results provided in sections III and IV, some simulation results are presented. We focus on the problem of structured covariance matrix estimation for a mixture noise.

In order to compare all estimates, we will plot the Mean Square Error (MSE) against the number  $K$  of data in the Gaussian case. The MSE used in this section is the following criterion:

$$MSE(\hat{\mathbf{M}}, \mathbf{M}) = E \left[ \frac{\|\hat{\mathbf{M}} - \mathbf{M}\|}{\|\mathbf{M}\|} \right],$$

where  $\|\cdot\|$  stands for the Frobenius norm and  $\hat{\mathbf{M}}$  denotes the studied estimate of  $\mathbf{M}$ .

The covariance matrix  $\mathbf{M}$  which has to be estimated is Toeplitz and is defined as follows

$$M_{ij} = \rho^{|i-j|} \quad \text{and} \quad N_{ij} = \tau^2 \rho^{|i-j|},$$

for  $1 \leq i, j \leq p$  and for  $0 < \rho < 1$ .

The size of the data is  $p = 3$  and the shape parameter  $\tau$  is equal to 1.

Several well-known covariance matrix estimates are compared to those provided by the EL method. This allows to evaluate performance of our method. The chosen estimates of  $\mathbf{N}$  are the following:

- The well-known Sample Covariance Matrix which corresponds to the ML estimate of the covariance matrix under Gaussian assumptions and defined as follows

$$\hat{\mathbf{N}}_{SCM} = \frac{1}{K} \sum_{k=1}^K \mathbf{c}_k \mathbf{c}_k^H.$$

$\hat{\mathbf{N}}_{SCM}$  is used as a benchmark but it is not appropriate to our problem since it does not take into account the structure of the real covariance matrix.

- To fill this gap, we use an appropriate estimate for Toeplitz matrix, first introduced in [2] and defined by

$$\hat{\mathbf{N}}_{B2} = \max_{\hat{\mathbf{N}} \in \mathcal{M}_2} \left( -\ln[\det(\hat{\mathbf{N}})] - \text{Tr}(\hat{\mathbf{N}}^{-1} \hat{\mathbf{N}}_{SCM}) \right),$$

where  $\mathcal{M}_2$  denotes the set of Toeplitz matrix:  $\mathcal{M}_2 = \{\mathbf{A} \in \mathbb{C}^{p \times p} | \text{for } i \neq j, A_{ij} = a_{|i-j|}\}$ .

- Moreover, since the EL estimate uses the third constraint, we also build the particular Burg estimate defined by

$$\hat{\mathbf{N}}_{B3} = \max_{\hat{\mathbf{N}} \in \mathcal{M}_3} \left( -\ln[\det(\hat{\mathbf{N}})] - \text{Tr}(\hat{\mathbf{N}}^{-1} \hat{\mathbf{N}}_{SCM}) \right),$$

with  $\mathcal{M}_3 = \{\mathbf{A} \in \mathbb{C}^{p \times p} | \text{for } \rho \in ]0, 1[, A_{ij} = \rho^{|i-j|}\}$ .

- Finally, EL estimates will be also compared to an estimate adapted to non-Gaussian noise. We choose the Fixed Point estimate [13] defined as

$$\hat{\mathbf{M}}_{FP} = \frac{m}{K} \sum_{k=1}^K \frac{\mathbf{c}_k \mathbf{c}_k^H}{\mathbf{c}_k^H \hat{\mathbf{M}}_{FP}^{-1} \mathbf{c}_k}.$$

Notice that  $\hat{\mathbf{M}}_{FP}$  is self-normalized and it does not depend on  $\tau$ . Thus, it provides directly an estimate of  $\mathbf{M}$ . For the other estimates, a normalization by an estimate of  $\tau^2$  has to be made. As  $\text{Tr}(\mathbf{M}) = p$ , one has  $\text{Tr}(\mathbf{N}) = \tau^2 p$  and thus, for all estimates, except  $\hat{\mathbf{M}}_{FP}$ , one has

$$\hat{\mathbf{M}} = \frac{p}{\text{Tr}(\hat{\mathbf{N}})} \hat{\mathbf{N}}.$$

Concerning the EL method, notations of section IV are still valid:  $\hat{\mathbf{N}}_{EL1}$ ,  $\hat{\mathbf{N}}_{EL2}$  and  $\hat{\mathbf{N}}_{EL3}$ .

To compare the different estimates in a non-Gaussian context, we retain three mixture scenarios for the data PDF: Weibull plus Gaussian, Student-tplus Gaussian and K-distribution plus Gaussian. These PDF are widely used in radar processing, see for example [15], [4], [7], [13]. This kind of noise modeling allows to obtain impulsive clutter

which is modulated by a Gaussian process.

Since the FP estimate, like the SCM and EL1, does not assume any particular structure on the covariance matrix  $\mathbf{M}$ , it makes no sense to compare these estimates to the others. Therefore, EL3 and B3 have to be considered together while EL4 and B4 are considered as a third set of estimates.

Since a major part of figures comments is available for the three situations. Figure 1, 2 and 3 are commented together at the end of section V.

#### A. Mixture of Weibull and Gaussian distributions

For  $k = 1, \dots, K$ ,

$$\mathbf{c}_k \sim \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}^{1/2} \frac{(1-\alpha)\mathcal{CW} + \frac{\alpha}{20}\mathcal{CN}(\mathbf{0}, \mathbf{I})}{\sqrt{(1-\alpha)^2 + (\frac{\alpha}{20})^2}}$$

where  $\mathcal{CW}$  is a complex Weibull distribution with parameters 1 and 0.4, which has been centered and normalized. The choice of coefficients has been designed to obtain impulsive data for small values of  $\alpha$ .

The set of parameters in this subsection is  $p = 3$ ,  $K = 100$  and  $\rho = 0.4$ . This value of the correlation coefficient allows to analyze the Burg's method in a suitable context.

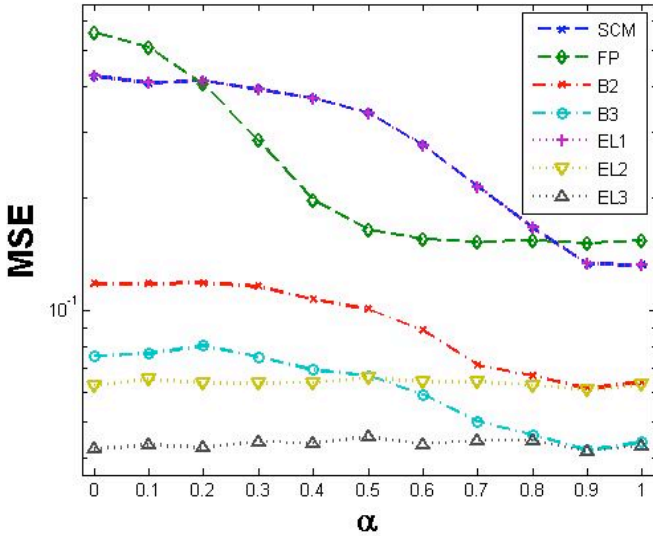


Fig. 1. MSE against  $\alpha$ , for a mixture of Weibull and Gaussian distributions, for  $\rho = 0.4$ ,  $p = 3$  and  $K = 100$ .

#### B. Mixture of Student-t and Gaussian distributions

For  $k = 1, \dots, K$ ,

$$\mathbf{c}_k \sim \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}^{1/2} \frac{(1-\alpha)\mathcal{CT} + \alpha\mathcal{CN}(\mathbf{0}, \mathbf{I})}{\sqrt{(1-\alpha)^2 + \alpha^2}}$$

where  $\mathcal{CT}$  is a complex Student-t distribution with degree of freedom 2.1, which has been centered and normalized.

The set of parameters in this subsection is  $p = 3$ ,  $K = 100$  and  $\rho = 0.5$ .

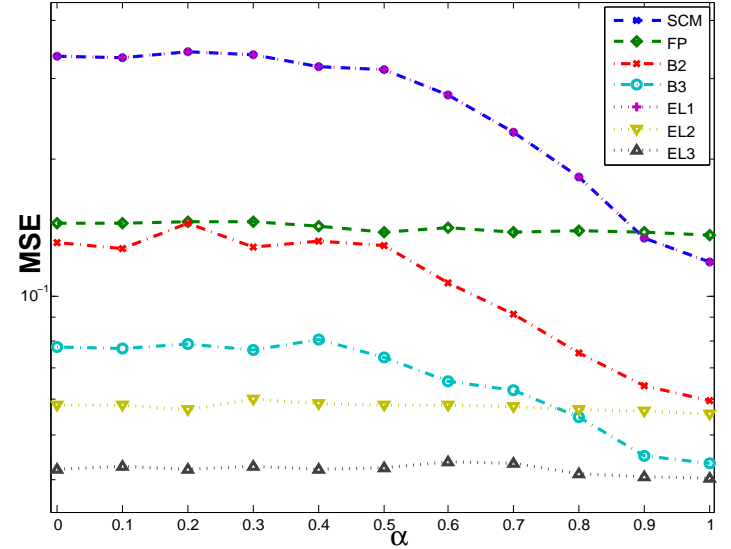


Fig. 2. MSE against  $\alpha$ , for a mixture of Student-t and Gaussian distributions, for  $\rho = 0.5$ ,  $p = 3$  and  $K = 100$ .

#### C. Mixture of K-Distribution and Gaussian distribution

For  $k = 1, \dots, K$ ,

$$\mathbf{c}_k \sim \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}^{1/2} \frac{(1-\alpha)\mathcal{CK}(\nu, \mathbf{0}, \mathbf{I}) + \alpha\mathcal{CN}(\mathbf{0}, \mathbf{I})}{\sqrt{(1-\alpha)^2 + \alpha^2}}$$

The K-distribution is the product of the square root of a random variable, Gamma distributed, and an independent complex zero mean Gaussian vector, with covariance matrix  $\mathbf{M}$ ; it is denoted  $\mathcal{CK}(\nu, \mathbf{0}, \mathbf{M})$ . The shape parameter  $\nu$  of the K-Distribution has been fixed to 0.01, this allows to generate very impulsive noise for small values of  $\alpha$ .

Again, the set of parameters in this subsection is  $p = 3$ ,  $K = 100$  and  $\rho = 0.5$ .

#### D. Simulations synthesis

A first comment is that, as expected, the MSE decrease as the estimates make use of more information. On each graphic, the estimates that do not use any prior on the structure (EL1 and SCM which are equal) have MSE higher than the MSE of the estimates that make use of the Toeplitz structure (EL2 and B2). These last are themselves higher than the MSE of the estimates using all the structure (EL3 and B3). FP has a specific behavior since it is designed to be robust against impulsive noises.

Notice that the ratio between the curves  $EL2$  and  $EL3$  (resp.  $B2$  and  $B3$ ) is constant for each value of  $\alpha$ . This ratio

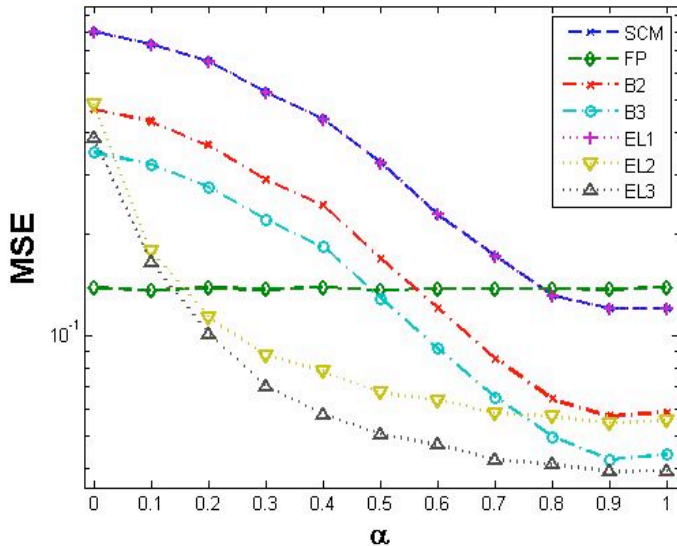


Fig. 3. MSE against  $\alpha$ , for a mixture of K-distribution and Gaussian distributions, for  $\rho = 0.5$ ,  $p = 3$  and  $K = 100$ .

corresponds theoretically to a ratio between the expectation of a  $\chi^2(3)$  (where 3 is the number of unknown parameters:  $\tau$ ,  $a_1$  and  $a_2$ ) and the expectation of a  $\chi^2(2)$  (where 2 is the number of unknown parameters:  $\tau$  and  $\rho$ ) and is equal to 1.5. This remark is illustrated on the three figures.

First, we comment the estimates which do not use any particular structure on the covariance matrix  $\mathbf{M}$ . EL1 and SCM curves decrease as  $\alpha$  increases, i.e. as the noise gets Gaussian. On the opposite, one can notice that in the cases of Student-t and K-Distribution noises, FP's MSE is constant for all values of  $\alpha$ , and is lower than EL1's MSE. Nevertheless, FP's estimate suffers some loss in performance for small values of  $\alpha$  in the Weibull case.

Then, we consider the estimate based on the Toeplitz structure of the matrix: EL2 and B2. On the three figures, EL2 performs better than B2, especially for small values of  $\alpha$ , i.e. in non-Gaussian noise. This was expected: Burg estimates, as the SCM, are based on the Gaussian assumption that does not hold for  $\alpha < 1$ . For  $\alpha = 1$ , EL2 and B2 provide approximately the same performance.

Finally, EL3 and B3 can be compared as they are based on the same special Toeplitz structure of the correlation matrix. As explained before, these two curves are parallel to EL2 and B2's and thus, the comments remain the same.

As a conclusion to this simulation section, one can notice that for the three scenarios, the methods based on EL perform better than the methods based on the classical Gaussian likelihood and have comparable behavior even when the data are actually Gaussian. Moreover, the EL method is robust to

the distribution variations since it is not based on an a priori distribution.

## VI. CONCLUSION

In this paper, we make use of a semi parametric method, the Empirical Likelihood, to estimate the covariance matrix of the clutter without using a model for its distribution. This method can therefore be applied whatever the distribution of the data and can prevent from the bias due to a model misspecification. An additional enjoyable property of the Empirical Likelihood is that priors can be easily taken into account in order to improve the performance of the estimate.

Moreover, in some cases, a closed-form expression of the estimate is derived, like for example for the classical Toeplitz structure estimation. This allows to avoid an optimization procedure which can be difficult to implement.

Finally, on simulations, it appears that the closed-form Empirical Likelihood estimate performs better than classical methods and adapts automatically to the data distribution. The performance are almost constant whatever the model and the distance to the Gaussian. Thus, the EL method is well-fitted to problems of clutter transitions and/or impulsive clutter of unknown distribution.

## REFERENCES

- [1] J B Billingsley. Ground clutter measurements for surface-sited radar. Technical Report 780, MIT, February 1993.
- [2] J P Burg, D G Luenberger, and D L Wenger. Estimation of structured covariance matrices. *Proc. IEEE*, 70(9):963–974, September 1982.
- [3] E Conte, M Lops, and G Ricci. Asymptotically optimum radar detection in compound-gaussian clutter. *IEEE Trans.-AES*, 31(2):617–625, April 1995.
- [4] E Conte, M Lops, and G Ricci. Adaptive detection schemes in compound-gaussian clutter. *IEEE Trans.-AES*, 34(4):1058–1069, October 1998.
- [5] D R Fuhrmann. Application of toeplitz covariance estimation to adaptive beamforming and detection. *IEEE Trans.-SP*, 39(10):2194–2198, October 1991.
- [6] F Gini. Sub-optimum coherent radar detection in a mixture of k-distributed and gaussian clutter. *IEE Proc. Radar, Sonar and Navigation*, 144(1):39–48, February 1997.
- [7] F Gini and M V Greco. Sub-optimum approach to adaptive coherent radar detection in compound-gaussian clutter. *IEEE Trans.-AES*, 35(3):1095–1103, July 1999.
- [8] 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*, 82(12):1847–1859, December 2002.
- [9] R M Gray. Toeplitz and circulant matrices. Technical Report 6504-1, Stanford University Information Systems Laboratory, April 1977.
- [10] U Grenander and G Szego. *Toeplitz Forms and Their Applications*. University of California Press, Berkeley and Los Angeles, 1958.
- [11] A B Owen. Empirical likelihood ratio confidence regions. *Annals of Statistics*, 18:90–120, 1990.
- [12] A B Owen. *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton, 2001.
- [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*, (to appear).
- [14] D S G Pollock. *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, 1999.
- [15] S Watts. Radar detection prediction in sea clutter using the compound k-distribution model. *IEE Proceeding, Part. F*, 132(7):613–620, December 1985.