The Empirical Likelihood method applied to covariance matrix estimation

Fréderic Pascal, Hugo Harari-Kermadec, Pascal Larzabal

The Empirical Likelihood method applied to Covariance Matrix Estimation

Frederic Pascal, Hugo Harari-Kermadec and Pascal Larzabal

Abstract

This paper presents a new estimation scheme for signal processing problems in unknown noise field. The Empirical Likelihood has been introduced in the mathematical community, but, surprisingly, it is still unknown in the Signal Processing community. This estimation method is an alternative to estimate unknown parameters without using a model for the Probability Density Function.

The aim of this paper is twofold: first, the Empirical Likelihood theory is presented and revisited thanks to the Moment Method. Its properties are derived. Secondly, to emphasize all the potentiality of this method, we address the problem of Toeplitz matrix estimation: this leads us to obtain improved estimates in comparison to conventional ones, as shown in simulations.

Key words: Empirical Likelihood, Maximum Likelihood, structured parameters estimation, non-Gaussian noise, covariance matrix estimation.

1 Introduction

It is often assumed that signals, interferences or noises are Gaussian stochastic processes. Indeed, this assumption makes sense in many applications. Among them, we can cite : sources localization in passive sonar, radar detection where thermal noise and clutter are often modeled as Gaussian processes, digital communications where the Gaussian hypothesis is widely used for interferences and noises.

In these contexts, Gaussian models have been thoroughly investigated in the framework of Statistical Estimation and Detection Theory [1,2,3]. They have led to attractive algorithms. For instance, we can cite the stochastic Maximum Likelihood method for sources localization in array processing [4,5], the matched filter and its adaptive variants in radar detection [6,7] and in digital communications [8].

However, such widespread techniques suffer from several drawbacks when the noise process is a non-Gaussian stochastic process [9]. Therefore, non-Gaussian noise modeling has gained many interest in the last decades and presently leads to active researches in the literature. Higher order moment methods [10] have initiated this research activity while particle filtering [11] is now intensively investigated. In radar applications, experimental clutter measurements, performed by MIT [12], showed that these data are not correctly described by Gaussian statistical models. More generally, numerous non-Gaussian models have been developed in several engineering fields [13,14,15].

Nevertheless, the question of a model choice for previous applications remains since, most of the time, chosen modeling does not perfectly describe the data behavior. And in these cases, classical estimation methods like for example Maximum Likelihood (ML) based on the data Probability Density Function (pdf) are used, leading as expected to only sub-optimal results. Several non-parametric techniques are proposed in the literature to estimate this unknown pdf. We can cite for example wavelet methods which have been widely investigated. But, most of them are difficult to implement.

An alternative is the Empirical Likelihood (EL) [16]. This method allows to estimate the unknown parameters without assuming a noise modeling. Moreover, prior informations on the data (known moments, parameter structure, ...) can be integrated in the processing by means of constraints in the optimization procedure. However, surprisingly, this estimation scheme is still unused in the area of Signal Processing estimation. To the best of our knowledge, we can only cite [17] and [18,19] in the corresponding literature.

The main interest of this method is to estimate unknown parameters of interest without assumption on the pdf. Furthermore, we will estimate only parameters of interest but we do not aim in estimating the pdf. This approach differs from robustness considerations, exposed for example by [20], that lead to M-estimators and minimax theory. Here, we do not have in mind a specific model, from which the true pdf could be slightly apart. However, EL will be shown to be based on a moment equation which links the parameters of interest to the observed data, like in the method of moments [1]. Moreover, EL relies on the optimization of a quantity that reminds of a likelihood.

The mathematical properties of EL have already been established in the statistical literature. The aim of this paper is to introduce the EL method to the signal processing community, by means of classical estimation problems. To illustrate the potentiality of the EL method, we analyze the problem of structured covariance matrix estimation. In particular, we focus on the persymmetric structure [21] and on the Toeplitz structure which is widely used [22,23,24,25,26].

Many applications rely on the covariance matrix estimation such as sources localization or Doppler frequencies detection. In these cases, performance directly depends on the accuracy of the covariance matrix estimate, see e.g. [27] for the particular case of detection in non Gaussian noise, or more generally, [5] in array processing problems. The choice of the covariance matrix estimation method affects dramatically the accuracy of the estimation in practical settings. In the field of channel identification, [18] compares EL with conventional methods.

The paper is organized as follows. Section 2 presents the estimation problem of interest while Section 3 gives an original presentation of the EL procedure, adapted to Signal Processing problems. Sections 3.4 and 4 present two applications of the EL method, first without constraint and then, the EL method uses prior informations. In these sections, comparison with the classical ML method will be analyzed through the problem of covariance matrix estimation under Gaussian assumptions. Then, Section 5 contains simulations which illustrate theoretical results of Section 4.

2 Problem Formulation

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

2.1 Notations

In the following, H denotes the conjugate transpose operator, $\top$ denotes the transpose operator, * denotes the conjugate operator, $E_P[f(\mathbf{x})]$ is the expectation of the function $f(\mathbf{x})$ when the random variable $\mathbf{x}$ is distributed according to the probability P, $E[\mathbf{x}|f(\mathbf{x})]$ means the expectation of $\mathbf{x}$ subject to $f(\mathbf{x})$. $\mathrm{Tr}(\mathbf{M})$ is the trace of matrix $\mathbf{M}$ and $\det(\mathbf{M})$ is the determinant 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 $\mathcal{R}e(z)$ and $\mathcal{I}m(z)$ its real and imaginary parts.

2.2 Statistical Framework

In a lot of signal processing problems, we have to extract the estimator $\widehat{\boldsymbol{\theta}}$ of parameters $\boldsymbol{\theta}$ based on some noisy data $\mathbf{x}$. This leads to the functional relation:

$$\widehat{\boldsymbol{\theta}} = T(\mathbf{x}). \tag{1}$$

To obtain a useful estimator $T(.)$ of parameter $\boldsymbol{\theta}$, a mathematical model on the data has to be introduced. One of the most widespread model, in the area of signal processing, is the following:

$$\mathbf{x}_k = \mathbf{h}(\boldsymbol{\theta}, k) + \mathbf{n}_k\,, \quad \text{for } k = 1, \ldots, K\,, \tag{2}$$

where $\mathbf{x}_k$ is the k^{th} observation data, $\mathbf{n}_k$ is an additive noise and $\mathbf{h}(\boldsymbol{\theta}, k)$ is the noiseless part of observation which depends on $\boldsymbol{\theta}$ and k. Sometimes, one has the separability property, $\mathbf{h}(\boldsymbol{\theta}, k) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_k$, where $\mathbf{A}(\boldsymbol{\theta})$ is a transfer matrix and $\mathbf{s}_k$ is a signal.

Recently, the huge increase of computer capacity has allowed the implementation of very sophisticated and performing methods. One of the most famous of them is the ML method which requires, by construction, the knowledge of the data pdf, up to parameters of interest and possibly, additional nuisance parameters. If such a pdf is given, ML methods lead to very performing results, optimal under specific conditions.

The performance of such estimators critically depends on the pdf assumption. This assumption is chosen to be consistent with the problem, but also to be mathematically convenient. Therefore, the choice is restricted to a small number of classical pdf distributions and the real data pdf can apart, at least slightly, from these families. It would be interesting to develop performing methods which do not require an assumption on the pdf family. In this paper, we propose to consider the Empirical Likelihood method, which is not based on the choice of a model for the pdf.

More precisely, let P_0 be the unknown distribution from which data are generated. In practice, this probability is unknown and a classical approach is to choose a parametric family for the pdf and to assume that the pdf corresponding to P_0 belongs to this family. The expression "parametric family" means that the pdf is known up to an element of a finite dimensional parameter space. For instance, the most currently used family is the set of Gaussian pdf, identified with two distinct parameters: the expectation and the covariance matrix (or variance in dimension 1). Notice that assuming a distribution family is a very restricting assumption since it reduces an infinite dimensional estimation problem to the estimation of finite dimensional parameters. EL method [16] has been designed as a means of relaxing the restrictions on the distribution P_0

in comparison to parametric approaches. Indeed, the only assumption is that the variance is finite. The aim of this method is to prevent from degradations due to model misspecifications. Therefore, instead of restricting the approach to the choice of a parametric family to mimic the data and then estimate the parameters, we look for the bigger flexibility.

3 Empirical likelihood

For a better understanding, this section develops the interest and the aim of using the Empirical Likelihood (EL) method, keeping in mind signal processing applications.

3.1 An extended method of moments

The method of moments is a classical method for estimation problems in which the parameter of interest $\boldsymbol{\theta}$ is defined as the solution of an equation involving the moments of the distribution [1]. This equation can be formulated as follows:

for some regular function $\mathbf{m}$ defined by

$$\mathbf{m} : \begin{cases} \mathbb{R}^p \times \mathbb{R}^d \longrightarrow \mathbb{R}^n \\ \mathbf{x}, \boldsymbol{\theta} \qquad \longrightarrow \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}) \end{cases}, \tag{3}$$

$\boldsymbol{\theta}_0$, the true value of the parameter, is the solution of

$$E_{P_0}[\mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_0)] = \mathbf{0}\,, \tag{4}$$

where $\mathbf{0}$ denotes the null vector with appropriate dimension (n here) and $n \geq d$, p is the dimension of the observations $\mathbf{x}$ and d the dimension of the parameter of interest $\boldsymbol{\theta}$. Notice that the expectation is taken under the true distribution P_0.

This kind of equation is called "moment condition" because it generalizes the equations giving the moments. Following equations illustrate the case of first and second moments respectively with unknown expectation μ and/or unknown covariance matrix Σ

$$E_{P_0}[\mathbf{x} - \mu] = \mathbf{0} \qquad \text{and} \qquad E_{P_0}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^H - \Sigma] = \mathbf{0}\,. \tag{5}$$

The method of moments [1] consists in solving in $\boldsymbol{\theta}$ the empirical version of

the equation (4):

$$\frac{1}{K}\sum_{k=1}^{K}\mathbf{m}(\mathbf{x}_k,\boldsymbol{\theta})=\mathbf{0}\,, \tag{6}$$

and $\widehat{\boldsymbol{\theta}}$, the method of moments estimator, will be the solution of equation (6). Formally, this can be written as follows

$$E_{P_K}[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})]=\mathbf{0}\,, \tag{7}$$

where

$$P_K(\mathbf{x})=\frac{1}{K}\sum_{k=1}^{K}\delta_{\mathbf{x}_k}(\mathbf{x})\,, \tag{8}$$

where $\delta_{\mathbf{X}}$ is the Dirac measure at element $\mathbf{x}$.

This very classical procedure gives goods results when the number K of data is large enough because this method relies on asymptotic considerations. But when the number of data is too small, the probability P_K may be a poor approximation of the true probability P_0, and as a consequence, the method of moments estimator will be far from the true value $\boldsymbol{\theta}_0$ of the parameter. To fill this gap, it is interesting to introduce the family of multinomial distribution G supported by the data set, which is defined as

$$G(\mathbf{x})=\sum_{k=1}^{K}q_k\,\delta_{\mathbf{x}_k}(\mathbf{x})\,, \tag{9}$$

where $0<q_k<1$ and $\sum_{k=1}^{K}q_k=1$. Notice that this is the identifiable choice of maximal dimension for a distribution family because it has as many degrees of freedom as they are observations. By substituting P_K by G, the moment equation (7) becomes

$$E_G[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})]=\mathbf{0}, \tag{10}$$

which rewrites

$$\sum_{k=1}^{K}q_k\mathbf{m}(\mathbf{x}_k,\boldsymbol{\theta})=\mathbf{0}\,, \tag{11}$$

In the estimation of parameter $\boldsymbol{\theta}$ of interest, the weights $(q_1,\ldots,q_K)$, introduced in equation (9) (which are the parameters of the multinomial) are nuisance parameters.

To obtain a concentrated criterion in the parameter of interest, we first optimize on G by using the Kullback-Leibler discrepancy $\mathcal{K}$, see Appendix A for details :

$$\mathcal{C}(\boldsymbol{\theta})=2K\min_{G}\left\{\mathcal{K}(G,P_K)\,\middle|\,E_G[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})]=\mathbf{0}\right\}. \tag{12}$$

This leads to the following estimator

$$\hat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}\{\mathcal{C}(\boldsymbol{\theta})\} \tag{13}$$

In fact, $\mathcal{C}(\boldsymbol{\theta})$ is the Empirical Likelihood Ratio (ELR), as we will see in the sequel by emphasizing the link between this method and the classical Maximum Likelihood method.

3.2 Likelihood, Empirical Likelihood and Empirical Likelihood Ratio

In this section, we show how the previous procedure can be understood in the classical likelihood context and how the criterion $\mathcal{C}$ can be interpreted as a likelihood ratio. We consider the multinomial G as a parametric model for the data. Notice that this parametric model assumption is only considered to formally introduce the EL method as a classical likelihood. For that purpose, we define, for G and $\boldsymbol{\theta}$ verifying the moment equation (10), $E_G[\mathbf{m}(\mathbf{x}, \boldsymbol{\theta})] = \mathbf{0}$, the pdf corresponding to G as follows

$$g_{(\boldsymbol{\theta},q_1,\ldots,q_K)}(\mathbf{x}) = \begin{cases} q_k & \text{if } \exists k,\ \mathbf{x} = \mathbf{x}_k, \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

To simplify the exposition, we suppose that there are no ties in the data. This issue can be easily handled, see [28], page 11. The likelihood function corresponding to the pdf $g_{(\boldsymbol{\theta},q_1,\ldots,q_K)}$, such as the moment equation (10) is verified, is called Empirical Likelihood and is defined as follows:

$$EL(\boldsymbol{\theta}) = \sup_{(q_1,\ldots,q_K)} \left\{ \prod_{k=1}^{K} g_{(\boldsymbol{\theta},q_1,\ldots,q_K)}(\mathbf{x}_k) \middle| E_G[\mathbf{m}(\mathbf{x}, \boldsymbol{\theta})] = \mathbf{0} \right\} \tag{15}$$

$$= \sup_{(q_1,\ldots,q_K)} \left\{ \prod_{k=1}^{K} q_k \middle| \sum_{k=1}^{K} q_k\, \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}) = \mathbf{0},\ \sum_{k=1}^{K} q_k = 1 \right\}. \tag{16}$$

In a similar way as in the classical ML theory, the log-likelihood ratio corresponding to $EL(\boldsymbol{\theta})$ is defined by:

$$ELR(\boldsymbol{\theta}) = -2 \log \left(\frac{EL(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}} \{EL(\boldsymbol{\theta})\}} \right). \tag{17}$$

The following theorem allows to make the link between $ELR(\boldsymbol{\theta})$ and $\mathcal{C}(\boldsymbol{\theta})$.

Theorem 3.1 (Connection between the criterion $\mathcal{C}(\boldsymbol{\theta})$ and $ELR(\boldsymbol{\theta})$) *If $n = d$, $\mathcal{C}(\boldsymbol{\theta})$, defined by equation (12), is equal to $ELR(\boldsymbol{\theta})$:*

$$\mathcal{C}(\boldsymbol{\theta}) = ELR(\boldsymbol{\theta}). \tag{18}$$

***Proof* 3.2.1**
For the clarity of the presentation, proof of Theorem 3.1 is postponed to Appendix B.

The main result of this Theorem is that the maximization of $\mathcal{C}(\boldsymbol{\theta})$ and $ELR(\boldsymbol{\theta})$ lead to the same estimator for $\boldsymbol{\theta}$. When $n \neq d$, this result is preserved since $\mathcal{C}(\boldsymbol{\theta})$ and $ELR(\boldsymbol{\theta})$ are not any more equal but proportional.

Thus, the estimator $\hat{\boldsymbol{\theta}}$ defined by equation (13) is the Maximum Empirical Likelihood (MEL) estimator $\hat{\boldsymbol{\theta}}_{MEL}$. The following theorem gives its asymptotic behavior.

Theorem 3.2 (Maximum Empirical Likelihood (MEL) estimator) *Let $\mathbf{x}_1, \ldots, \mathbf{x}_K$ be an i.i.d. sample distributed according to a distribution P_0, such as $E_{P_0}[\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}_0)] = \mathbf{0}$. If $\mathbf{D} = E_{P_0}[\partial \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}]$ has rank n and $\mathbf{M} = E_{P_0}[\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}_0)\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}_0)^\top]$ is definite positive then the MEL estimator, equivalently given by*

$$\hat{\boldsymbol{\theta}}_{MEL} = \arg\max_{\boldsymbol{\theta}} \{EL(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \{ELR(\boldsymbol{\theta})\} \tag{19}$$

is an asymptotically normal estimator of $\boldsymbol{\theta}_0$:

$$\sqrt{K}\left(\hat{\boldsymbol{\theta}}_{MEL} - \boldsymbol{\theta}_0\right) \xrightarrow[K\to\infty]{dist.} \mathcal{N}\left(\mathbf{0}, (\mathbf{D}^\top \mathbf{M} \mathbf{D})^{-1}\right) . \tag{20}$$

***Proof* 3.2.2**
See [29].

Theorem 3.2 gives asymptotic statistical behavior of the MEL estimator in a general context. The MEL estimator is so defined as the solution of an optimization equation (see equation (19)), resolved thanks to a Lagrangian method. It is possible to add some prior information, on the parameter structure or on known moments of the data, into the optimization equation. This could be written as follows:

$$\hat{\boldsymbol{\theta}}_{MEL} = \arg\max_{\boldsymbol{\theta}} \{\, EL(\boldsymbol{\theta}) |\, \boldsymbol{\theta} \in \mathcal{E}\} \ , \tag{21}$$

where $\mathcal{E}$ is the set of constraints on the parameter $\boldsymbol{\theta}$ of interest. Section 4 presents an estimation problem in which some prior informations on the parameter of interest are gradually integrated in the EL procedure. Notice that the prior informations may concern the parameter structure, the introduced parameterization $\mathcal{P}(\boldsymbol{\theta})$ and the known moments of the observations.

In several situations, a closed-form expression for the MEL estimator can be obtained. This is the case when the parameter structure or the parameterization can be expressed in a linear form. This is the purpose of the following section.

Some particular problems lead to closed-form expressions for the MEL estimator. This subsection presents the case where the moment function $\mathbf{m}$ is a linear function, with $n = p$. This is the case when the parameter of interest is an expectation and that some additional prior information is available. The estimation scheme can take into account this prior information by introducing an augmented observation vector as proposed in the following. For notational convenience, this augmented vector will be also denoted $\mathbf{x}$.

$\mathbf{x}$ is divided in two parts: $\mathbf{y}$, a first data transformation traducing the parameter estimation problem such that $E_{P_0}[\mathbf{y}] = \boldsymbol{\theta}$, and $\mathbf{z}$ a second data transformation reflecting the prior information. This procedure leads to

$$\mathbf{x} = \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}, \text{ where } E_{P_0}[\mathbf{z}] = \mathbf{0}. \tag{22}$$

Therefore, the function $\mathbf{m}$, defined by

$$\mathbf{m} : \begin{cases} (\mathbb{R}^p, \mathbb{R}^d) \to \mathbb{R}^p \\ (\mathbf{x}, \boldsymbol{\theta}) = \left(\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}, \boldsymbol{\theta} \right) \mapsto \begin{pmatrix} \mathbf{y} - \boldsymbol{\theta} \\ \mathbf{z} \end{pmatrix} \end{cases}, \tag{23}$$

is such that $E_{P_0}[\mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_0)] = \mathbf{0}$ and this leads to the following theorem.

Theorem 3.3 (Estimation with additional constraints)
Let $(\mathbf{x}_1, \ldots, \mathbf{x}_K)$ be an i.i.d. data set in $\mathbb{R}^p$, with common probability P_0 and for $n = p$. For $k = 1, \ldots, K$, let us set $\mathbf{x}_k = \begin{pmatrix} \mathbf{y}_k \\ \mathbf{z}_k \end{pmatrix} \in \mathbb{R}^d \times \mathbb{R}^{p-d}$ and assume that the expectation and variance-covariance matrix are

$$E_{P_0} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}_0 \\ \mathbf{0} \end{pmatrix} \qquad \text{and} \qquad Var \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} V_{yy} & V_{yz} \\ V_{zy} & V_{zz} \end{pmatrix}. \tag{24}$$

Then the Maximal Empirical Likelihood estimator is given by

$$\hat{\boldsymbol{\theta}}_{MEL} = \overline{\mathbf{y}} - V_{yz} V_{zz}^{-1} \overline{\mathbf{z}}\,. \tag{25}$$

where $\overline{\mathbf{y}}$ (resp. $\overline{\mathbf{z}}$) denotes the empirical mean, $\overline{\mathbf{y}} = \dfrac{1}{K} \displaystyle\sum_{k=1}^{K} \mathbf{y}_k$ (resp. $\overline{\mathbf{z}} = \dfrac{1}{K} \displaystyle\sum_{k=1}^{K} \mathbf{z}_k$).

***Proof* 3.3.1**
This result has been proved in [28], page 52.

In practice, V_{yz} and V_{zz} are replaced by their empirical versions that are consistent estimators.

Notice that when there is no additional information, i.e. $\mathbf{z} = \mathbf{0}$, one obtains the following result:

$$\hat{\boldsymbol{\theta}}_{MEL} = \overline{\mathbf{y}}\,. \tag{26}$$

This is also the case when $\mathbf{z}$ is uncorrelated with $\mathbf{y}$ because V_{yz} is the null matrix.

The following simple example illustrates the use of extended data vector. Let $\mathbf{x}$ be a 2-random vector of mean $\boldsymbol{\theta}$ and with identity covariance matrix $\mathbf{I}$. Then, the extended vector $\widetilde{\mathbf{x}}$ will be:

$$\widetilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathrm{vec}(\mathbf{x}\mathbf{x}^H - \mathbf{I}) \end{bmatrix}, \tag{27}$$

such that $E_{P_0}[\mathbf{x}] = \boldsymbol{\theta}$ and $E_{P_0}\left[\mathrm{vec}(\mathbf{x}\mathbf{x}^H - \mathbf{I})\right] = 0$ by definition of the problem. Thus, equation (25) provides a closed-form expression of the MEL estimator. This example deals with the prior information the second order moment of the observation data.

3.4 A simple example: mean estimation in presence of Gaussian noise

This section is devoted to the development of the EL procedure in a simple case. For that purpose, we first recap classical results obtained with the ML method under Gaussian assumptions. For that purpose, we focus on the following particular model:

$$\mathbf{x} = \boldsymbol{\theta}_0 + \mathbf{n}\,, \tag{28}$$

where $\mathbf{x}$ is the observation data, $\mathbf{n}$ is zero-mean and has known covariance matrix $\mathbf{M} = E[\mathbf{n}\mathbf{n}^H]$ and $\boldsymbol{\theta}_0$ is the unknown true value of the parameter of interest. As said previously, the EL method does not require the additive noise distribution knowledge but to build the classical likelihood, $\mathbf{n}$ will be assumed to be Gaussian: $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\theta}_0, \mathbf{M})$.

To estimate the parameter of interest $\boldsymbol{\theta}$, the natural procedure consists to build the profiled likelihood from an i.i.d. sample $(\mathbf{x}_1, \ldots, \mathbf{x}_K)$, defined as follows

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{K} \phi\left(\mathbf{M}^{-1/2}(\mathbf{x}_k - \boldsymbol{\theta})\right) \tag{29}$$

where ϕ stands for the standard Gaussian pdf in $\mathbb{R}^p$. This method has been extensively studied and provides under Gaussian assumptions the ML estimator $\hat{\boldsymbol{\theta}}_{ML}$, given by

$$\hat{\boldsymbol{\theta}}_{ML} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}_k = \overline{\mathbf{x}}. \tag{30}$$

Notice that the ML estimator of $\boldsymbol{\theta}_0$ defined at equation (30) does not depend on the true value of the scale parameter.

Let us now turn to the EL method. The natural moment equation corresponding to our example should simply be

$$E\left[\mathbf{x} - \boldsymbol{\theta}\right] = \mathbf{0}, \tag{31}$$

and then the moment function corresponding to this problem is

$$\mathbf{m}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x} - \boldsymbol{\theta}\,. \tag{32}$$

As seen previously in subsection 3.3, this implies that the MEL estimator is defined by

$$\hat{\boldsymbol{\theta}}_{MEL} = \overline{\mathbf{x}} = \hat{\boldsymbol{\theta}}_{ML}\,. \tag{33}$$

Remark 3.1
The conclusion of this section is that the estimator of the expectation provided by the MEL theory is the same as the ML estimator built under Gaussian assumptions (and under many classical others). This is an important result since problems with Gaussian models are extensively studied in Signal Processing and in this reference case, previous result (equation (33)) ensures that the MEL estimator performs exactly like the ML estimator.

4 Application to Covariance matrix estimation

In this section, we focus on the noise covariance matrix estimation problem. Let us set $\mathbf{x}$ a complex Gaussian p-vector with zero-mean and covariance matrix $E\left[\mathbf{x}\mathbf{x}^H\right] = \tau^2\,\mathbf{M}$, denoted $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \tau^2\mathbf{M})$. We set $\mathrm{Tr}(\mathbf{M}) = p$ for identifiability considerations. For simplicity matters, we set $\boldsymbol{\theta} = \tau^2\mathbf{M}$. Since $\boldsymbol{\theta}$ is Hermitian, we only need to estimate the upper triangular part of $\boldsymbol{\theta}$, which is our parameter of interest.

In the following, the problem statement will be modified according to prior informations assumed on the covariance matrix structure. The possible moment equations will be noted

$$E[\mathbf{m}_i(\mathbf{x}, \boldsymbol{\theta})] = \mathbf{0}. \tag{34}$$

In the sequel, d_i will stand for the number of unknown real parameters, i.e. the dimension of $\boldsymbol{\theta}$ and n_i for the dimension of $\mathbf{m}_i$. Notice that each moment

equation i will provide a different estimator. The different situations are the following:

(1) no prior information,
(2) the first moment is known, $\mathbf{x}$ is zero-mean,
(P) $\boldsymbol{\theta}$ is Persymmetric,
(3) the structure of $\boldsymbol{\theta}$ is Toeplitz,
(4) a prior information on the parameterization, $\boldsymbol{\theta}$ is parameterized with two unknown parameters, $\boldsymbol{\theta} = \mathcal{P}(\tau, \rho)$.

This allows to show the EL procedure development for different prior informations.

When there is no prior information on the covariance matrix structure, the moment equation is given by

$$E[\mathbf{m}_1(\mathbf{x}, \boldsymbol{\theta})] = E\left[\left(x_i x_j^H - \theta_{ij}\right)_{i \leq j}\right] = \mathbf{0}, \tag{35}$$

where x_i is the i^{th} element of the vector $\mathbf{x}$ and x^H stands for the conjugate of element x. We consider here $\boldsymbol{\theta}$ as an element of $\mathbb{C}^{p \times p}$, and then, the number of unknown real parameters is $d_1 = p^2$: p for the real valued elements θ_{jj} of the diagonal and $2p(p-1)/2$ for the complex valued elements θ_{ij}, $i < j$, of the strictly upper triangular part. The dimension n_1 of $\mathbf{m}_1$ is also equal to p^2 and then the MEL weights are $1/K$ (see Appendix B), and the corresponding estimator is

$$\widehat{\boldsymbol{\theta}}_{EL1} = \frac{1}{K} \sum_k \mathbf{x}_k \mathbf{x}_k^H = \overline{\mathbf{x}\mathbf{x}^H}\,. \tag{36}$$

The point of this section is to show how the estimator can be improved in the direction proposed by [29]: increase the information used by constraining the likelihood to fit prior knowledge, i.e. increase the dimension n of $\mathbf{m}$, or conversely decrease the dimension d of the parameter of interest $\boldsymbol{\theta}$.

Remark 4.1
An important property of EL is to easily take into account available prior information in the estimation process. Notice that in the problem under study in the current section, one has $E[\mathbf{x}] = \mathbf{0}$ so that it could be possible to use the more complete moment equation defined as follows

$$\mathbf{m}_2(\mathbf{x}, \boldsymbol{\theta}) = \begin{pmatrix} \left(x_i x_j^H - \theta_{ij}\right)_{i \leq j} \\ \mathbf{x} \end{pmatrix}, \text{ with } d_2 = p^2 \text{ and } n_2 = p(p+2) \tag{37}$$

This would increase the dimension n_2 of $\mathbf{m}_2$ to $n_1 + 2p = p(p+2)$ whereas the number d_2 of unknown parameters remains unchanged.

This estimator writes

$$\widehat{\boldsymbol{\theta}}_{EL2} = \arg\sup_{\boldsymbol{\theta}} \{EL_2(\boldsymbol{\theta})\} = \arg\min_{(\boldsymbol{\theta},\lambda)} \left\{ \sum_{k=1}^{K} \log\left(1 + \lambda^\top \mathbf{m}_2(\mathbf{x}_k, \boldsymbol{\theta})\right) \right\} . \quad (38)$$

Unfortunately, it is well known that first and second moment estimators are independent in the Gaussian case by the Student theorem. Therefore, the new constraint does not bring any supplementary information on the estimation of $\boldsymbol{\theta}$ *(if the data are Gaussian). The effect of the independence in the Gaussian case can be illustrate through Theorem 3.3: if* $\mathbf{y}$ *and* $\mathbf{z}$ *are asymptotically uncorrelated,* V_{yz} *is null and then the estimator is*

$$\hat{\boldsymbol{\theta}}_{MEL} = \overline{\mathbf{y}} - V_{yz} V_{zz}^{-1} \overline{\mathbf{z}} = \overline{\mathbf{y}}. \quad (39)$$

In the following, this constraint will not be used. However, under non-Gaussian assumptions, the zero-mean prior information could be useful for the covariance matrix estimation.

4.1 Prior information on the structure: θ *is an Hermitian Persymmetric matrix*

Many applications can result in a clutter covariance matrix that exhibits some particular structure. Such a situation is frequently met in radar systems using a symmetrically spaced linear array for spatial domain processing, or symmetrically spaced pulse train for temporal domain processing [30,31]. In these systems, the clutter covariance matrix $\boldsymbol{\theta}$ has the persymmetric property:

$$\boldsymbol{\theta} = \mathbf{J}_m \, \boldsymbol{\theta}^* \, \mathbf{J}_m \, , \quad (40)$$

where $\mathbf{J}_m$ is the m-dimensional antidiagonal matrix having 1 as non-zero elements and * stands here for the conjugate operator.

In this case, the moment equation becomes

$$E[\mathbf{m}_P(\mathbf{x}, \boldsymbol{\theta})] = E\left[\left(x_i x_j^H - \theta_{ij} \right)_{i \le j \le p+1-i} \right] = \mathbf{0}. \quad (41)$$

The number of unknown real parameters is $d_P = \dfrac{p(p+1)}{2}$: for p even: $p/2$ for the real valued elements θ_{jj} of the diagonal, p for the real valued elements ($p/2$ complex elements) $\theta_{j,p+1-j}$ of the anti-diagonal and $p(p-2)/2$ for the real valued of the strictly upper corner ($i < j < p+1-i$). The same dimension

d_P is obtained for odd p. The dimension of m is also equal to $\frac{p(p+1)}{2}$.

This leads to a new estimator $\widehat{\boldsymbol{\theta}}_{ELP}$ of $\boldsymbol{\theta}$, which integrates the constraint on the persymmetric structure of the real covariance matrix $\mathbf{M}$. This writes

$$\widehat{\boldsymbol{\theta}}_{ELP} = \arg\sup_{\boldsymbol{\theta}} \{EL_P(\boldsymbol{\theta})\} = \arg\min_{(\boldsymbol{\theta},\lambda)} \left\{ \sum_{k=1}^{K} \log\left(1 + \lambda^\top \mathbf{m}_P(\mathbf{x}_k, \boldsymbol{\theta})\right) \right\} . \quad (42)$$

We can rewrite the constraint in terms of expectations in order to obtain an explicit form of the estimator by means of Theorem 3.3 . For illustration, the constraints is written for $p = 3$ and for real covariance matrix. Let us define

$$\mathbf{y} = \left(\mathcal{R}e(x_1x_1^H, x_2x_2^H, x_1x_2^H, x_1x_3^H)\right)^\top , \quad (43)$$

$$\mathbf{z} = \left(x_1x_1^H - x_3x_3^H, x_1x_2^H - x_2x_3^H, \mathcal{I}m(x_1x_2^H, x_1x_3^H)\right)^\top . \quad (44)$$

Then, Theorem 3.3 gives an estimator of $\boldsymbol{\theta}$, thanks to previous quantities:

$$\widehat{\boldsymbol{\theta}}_{ELP} = \overline{\mathbf{y}} - V_{yz}V_{zz}^{-1}\overline{\mathbf{z}} \quad (45)$$

4.2 Prior information on the structure: θ is a Toeplitz matrix

A first step is to assume that $\boldsymbol{\theta}$ has real valued elements. The dimension of the parameter is therefore reduced to $p(p+1)/2$. The number of constraints remains unchanged because the constraints remain complex.

Several problems in Signal Processing assume that the covariance matrix of the additive noise has a Toeplitz structure [22,23,24,25]. 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. [26] is a good tutorial on Toeplitz matrices and contains most of their properties.

Notice that there already exists methods for structured covariance matrix estimation in which the Toeplitz case is treated in the ML framework, see e.g. [32,33]. We propose here to extend that to the MEL.

Now, let us assume that $\mathbf{M}$ is a Toeplitz matrix with trace p, $\mathbf{M}$ can be written

as follows

$$\exists (a_1, \ldots, a_{p-1}) \in \mathbb{R}^{p-1}, \begin{cases} M_{ij} = a_{|i-j|}, \text{ for } i \neq j\,, \\ M_{ii} = 1\,. \end{cases} \tag{46}$$

i.e.

$$\mathbf{M} = \mathbf{I} + \sum_{i=2}^{p} \left(a_{i-1}\mathbf{J}_i + a_{i-1}^H \mathbf{J}_i^\top \right), \tag{47}$$

where $\mathbf{I}$ stands for the identity matrix with appropriate dimension ($p \times p$ here) and $\mathbf{J}_i$ is the $p \times p$ matrix with 1's on the i^{th} upper diagonal and 0 everywhere else.

In this subsection, we take advantage of the Toeplitz structure on $\mathbf{M}$, and as a consequence on $\boldsymbol{\theta} = \tau^2\,\mathbf{M}$. Then, the constraint used in our process is the following: elements of the main diagonal θ_{ii} are all equal to τ^2 and elements on the upper diagonals θ_{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, $d_3 = p$. The moment conditions are modified to take the structure into account.

$$\mathbf{m}_3(\mathbf{x}, \boldsymbol{\theta}) = \begin{pmatrix} \left(x_i x_{i+j}^H - \tau^2 a_j \right)_{1 \leq i \leq p,\, 1 \leq j \leq p-i} \\ \left(x_i^2 - \tau^2 \right)_{1 \leq i \leq p} \end{pmatrix}, \text{ with } d_3 = p \text{ and } n_3 = \frac{p(p+1)}{2}. \tag{48}$$

This leads to a new estimator $\widehat{\boldsymbol{\theta}}_{EL3}$ of $\boldsymbol{\theta}$, which integrates the constraint on the Toeplitz structure of the real covariance matrix $\mathbf{M}$. This writes

$$\widehat{\boldsymbol{\theta}}_{EL3} = \arg\sup_{\boldsymbol{\theta}} \{EL_3(\boldsymbol{\theta})\} = \arg\min_{(\boldsymbol{\theta}, \lambda)} \left\{ \sum_{k=1}^{K} \log\left(1 + \lambda^\top \mathbf{m}_3(\mathbf{x}_k, \boldsymbol{\theta}) \right) \right\}. \tag{49}$$

We can rewrite the constraint in terms of expectations in order to obtain an explicit form of the estimator by means of Theorem 3.3 . For illustration, the constraints is written for $p = 3$. Define

$$\mathbf{y} = \mathcal{R}e\left(x_1 x_1^H, x_1 x_2^H, x_1 x_3^H \right)^\top \tag{50}$$

$$\mathbf{z} = \left(\mathcal{I}m\left(x_1 x_2^H, x_1 x_3^H \right), x_1 x_1^H - x_2 x_2^H, x_1 x_1^H - x_3 x_3^H, x_1 x_2^H - x_2 x_3^H \right)^\top \tag{51}$$

Then, Theorem 3.3 gives estimators for the first line of $\boldsymbol{\theta}$:

$$\left(\widehat{\tau^2}, \widehat{\tau^2 a_1}, \widehat{\tau^2 a_2} \right)^\top = \overline{\mathbf{y}} - V_{yz} V_{zz}^{-1} \overline{\mathbf{z}} \tag{52}$$

The estimator $\widehat{\boldsymbol{\theta}}_{EL3}$ writes then

$$\widehat{\boldsymbol{\theta}}_{EL3} = \begin{bmatrix} \widehat{\tau^2} & \widehat{\tau^2 a_1} & \widehat{\tau^2 a_2} \\ \widehat{\tau^2 a_1} & \widehat{\tau^2} & \widehat{\tau^2 a_1} \\ \widehat{\tau^2 a_2} & \widehat{\tau^2 a_1} & \widehat{\tau^2} \end{bmatrix} \tag{53}$$

4.3 Second prior information: θ is parameterized by τ and ρ

For some applications, a more structured matrix $\boldsymbol{\theta}(\tau, \rho)$ can be taken into account. The correlation information contained in the covariance matrix is assumed to be reduced to only one parameter, the correlation coefficient ρ:

$$M_{ij} = \rho^{|i-j|} \text{ and } \theta_{ij} = \tau^2 \rho^{|i-j|}, \tag{54}$$

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 ρ, which characterizes the correlation of the data.

Then, the unknown parameters of $\boldsymbol{\theta}$ are 2 real scalars, ρ and τ. Thus, the dimension of the problem is $d_4 = 2$, while the moment condition still remains unchanged, $n_4 = \frac{p(p+1)}{2}$:

$$\mathbf{m}_4(\mathbf{x}, \boldsymbol{\theta}) = \left(\left(x_i x_{i+j}^H - \tau^2 \rho^j \right)_{1 \leq i \leq p,\, 0 \leq j \leq p-i} \right), \text{ with } d_4 = 2 \text{ and } n_4 = \frac{p(p+1)}{2}. \tag{55}$$

This leads to the last estimator $\widehat{\boldsymbol{\theta}}_{EL4}$ of $\boldsymbol{\theta}$, defined by

$$\widehat{\boldsymbol{\theta}}_{EL4} = \arg\sup_{\boldsymbol{\theta}} \{EL_4(\boldsymbol{\theta})\} = \arg\min_{(\boldsymbol{\theta}, \lambda)} \left\{ \sum_{k=1}^{K} \log\left(1 + \lambda^\top \mathbf{m}_4(\mathbf{x}_k, \boldsymbol{\theta})\right) \right\}. \tag{56}$$

Remark 4.2
Notice that no closed-form expression of this last estimator is available, because we have not been able to write the constraints $x_i x_{i+j} - \tau^2 \rho^j$ as expectations in order to use Theorem 3.3. One can give a general expression which is also valid for the previous estimators:

$$\widehat{\boldsymbol{\theta}}_{ELj} = \sum_{k=1}^{K} q_k^*(j)\, \mathbf{x}_k \mathbf{x}_k^H, \tag{57}$$

where the $q_k^*(j)$ depends on the constraints (see Appendix B, Lemma B.1). For example, in the case of no constraint, i.e. $\widehat{\boldsymbol{\theta}}_{EL1}$, the $q_k^*(1)$ are, for $1 < k < K$, equal to $\dfrac{1}{K}$. This corresponds to the ML estimator of the covariance matrix for a Gaussian vector.

For the other estimators, the $q_k^*(j)$ allow to give a weight on the k^{th} data $\mathbf{x}_k$ in order to fulfill the a priori conditions. As the consequence, $\widehat{\boldsymbol{\theta}}_{EL3}$ has a Toeplitz structure while $\widehat{\boldsymbol{\theta}}_{EL4}$ satisfies the special Toeplitz structure given by ρ.

Notice that adding a priori information that can not be expressed as linear constraint leads to an increasing of the computational loads. In such a case, the computational loads and the estimation accuracy are in trade-off.

Remark 4.3
Taking into account linear constraints can also be easy with the conventional ML method. When using EL, adding a linear constraint leads to a better accuracy for the same computational complexity. The only additional cost is the simple estimation of the correlation and variance V_{yz} and V_{zz} in Theorem 3.3 . This estimation is accurate to the order $O(n^{-1})$ and is therefore negligible.

Nevertheless, non linear constraints on θ, such as the one considered in equation (55), is much more challenging for conventional methods and it requires each time the derivation of an appropriate algorithm. In opposite, EL is designed to handle any kind of additional constraint without any new investigation.

These theoretical estimators of $\boldsymbol{\theta}$ will be compared in the section 4 thanks to simulations on their Mean Square Error (MSE) and an expression for each one will be given.

5 Simulations

In order to enlighten results provided in sections 3.4 and 4, some simulation results are presented. We focus on the problem of structured covariance matrix estimation under Gaussian assumptions. Simulations are first performed with complex Gaussian noise and then with complex non-Gaussian noise. In this section, we only consider the true value of the parameters. For simplicity of the notation, the index 0 is omitted.

In order to compare all the previous estimators, 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:

$$MSE(\widehat{\mathbf{M}}, \mathbf{M}) = E\left[\frac{\|\widehat{\mathbf{M}} - \mathbf{M}\|}{\|\mathbf{M}\|}\right], \tag{58}$$

where $\|.\|$ stands for the Frobenius norm. Actually, this expectation is replaced

by its empirical form

$$\overline{MSE} = \frac{1}{L}\sum_{l=1}^{L} MSE(l)\,,$$

which converges to the MSE thanks to Strong Law of Large Numbers on the number L of Monte Carlo trials.

The covariance matrix $\boldsymbol{\theta} = \tau\,\mathbf{M}$ which has to be estimated has a particular Toeplitz structure and is defined by equation (54).

The size of the data is $p = 3$, the shape parameter τ is equal to 1. The correlation coefficient ρ is equal to 0.1 or 0.9. This allows to build a covariance matrix close to the identity matrix (i.e. $\rho = 0.1$) or a covariance matrix of very correlated data (i.e. $\rho = 0.9$). This choice of different ρ allows to test the robustness of covariance matrix estimation to data correlation. In particular, Burg's method [32], based on the inversion of the estimated matrix, is expected to suffer from the correlation. Therefore, for $\rho = 0.9$, it is expected a decrease in estimation performance.

For that purpose, several well-known covariance matrix estimators are compared to those provided by the EL method. This allows to evaluate performance of our method in comparison with classical ones. The chosen estimators of $\boldsymbol{\theta}$ are the following:

- The well-known Sample Covariance Matrix (SCM) which corresponds to the ML estimator of the covariance matrix under Gaussian assumptions and which is defined as follows

$$\widehat{\boldsymbol{\theta}}_{SCM} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^H\,. \tag{59}$$

  $\widehat{\boldsymbol{\theta}}_{SCM}$ 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 lack, we use an appropriate estimator for persymmetric matrix, derived from [32]:

$$\widehat{\boldsymbol{\theta}}_{B2} = \max_{\widehat{\boldsymbol{\theta}} \in \mathcal{M}_2} \left(-\ln[\det(\widehat{\boldsymbol{\theta}})] - \mathrm{Tr}(\widehat{\boldsymbol{\theta}}^{-1}\,\widehat{\boldsymbol{\theta}}_{SCM}) \right)\,, \tag{60}$$

  where $\mathcal{M}_2$ denotes the set of persymmetric matrix:
  $\mathcal{M}_2 = \{\mathbf{A} \in \mathbb{C}^{p\times p} | \mathbf{A} = \mathbf{J}_m\,\mathbf{A}^*\,\mathbf{J}_m\,,\}$, where $\mathbf{J}_m$ is the m-dimensional antidiagonal matrix having 1 as non-zero elements.
- To fill this lack, we use an appropriate estimator for Toeplitz matrix, first

introduced in [32] and defined by

$$\widehat{\boldsymbol{\theta}}_{B3} = \max_{\widehat{\boldsymbol{\theta}} \in \mathcal{M}_3} \left(-\ln[\det(\widehat{\boldsymbol{\theta}})] - \mathrm{Tr}(\widehat{\boldsymbol{\theta}}^{-1}\, \widehat{\boldsymbol{\theta}}_{SCM}) \right) , \tag{61}$$

where $\mathcal{M}_3$ denotes the set of Toeplitz matrix:
$\mathcal{M}_3 = \left\{ \mathbf{A} \in \mathbb{C}^{p \times p} | \text{for } i \neq j,\ A_{ij} = \tau^2 a_{|i-j|},\ \text{and } A_{ii} = \tau^2 \right\}$.

- Moreover, since the EL estimator uses the third constraint (i.e. when the only unknown parameter is the correlation coefficient ρ), we also build the particular Burg estimator defined by

$$\widehat{\boldsymbol{\theta}}_{B4} = \max_{\widehat{\boldsymbol{\theta}} \in \mathcal{M}_4} \left(-\ln[\det(\widehat{\boldsymbol{\theta}})] - \mathrm{Tr}(\widehat{\boldsymbol{\theta}}^{-1}\, \widehat{\boldsymbol{\theta}}_{SCM}) \right) , \tag{62}$$

where $\mathcal{M}_4 = \left\{ \mathbf{A} \in \mathbb{C}^{p \times p} | \text{for } 0 < \rho < 1,\ A_{ij} = \tau^2 \rho^{|i-j|} \right\}$.

- Finally, EL estimators will be also compared to a recently introduced estimator devoted to non-Gaussian noise, the Fixed Point estimator [34,35,36] defined as

$$\widehat{\mathbf{M}}_{FP} = \frac{m}{K} \sum_{k=1}^{K} \frac{\mathbf{x}_k \mathbf{x}_k^H}{\mathbf{x}_k^H \widehat{\mathbf{M}}_{FP}^{-1} \mathbf{x}_k} . \tag{63}$$

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

$$\widehat{\mathbf{M}} = \frac{p}{\mathrm{Tr}(\widehat{\boldsymbol{\theta}})} \widehat{\boldsymbol{\theta}} . \tag{64}$$

Concerning the EL method, notations of section 3.4 are still valid: $\widehat{\boldsymbol{\theta}}_{EL1}$ (which is equal to $\widehat{\boldsymbol{\theta}}_{EL2}$ and $\widehat{\boldsymbol{\theta}}_{SCM}$), $\widehat{\boldsymbol{\theta}}_{ELP}$, $\widehat{\boldsymbol{\theta}}_{EL3}$ and $\widehat{\boldsymbol{\theta}}_{EL4}$.

5.1 Gaussian case

Now we give the mean square errors (MSE) of the corresponding estimation procedures, for different values of the data set length K and for the seven estimators of interest when the data are Gaussian p-vector with zero-mean and the covariance matrix $\boldsymbol{\theta} = \tau^2 \mathbf{M}$, with $p = 3$ and $\tau = 1$, i.e.

$$\text{for } k = 1, \ldots, K\ ,\ \mathbf{x}_k \sim \mathcal{CN} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} , \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \right) \tag{65}$$

5.1.1 $\rho = 0.1$

Figure B.1 is plotted with a logarithmic scale on both horizontal and vertical axes.

The curves which correspond to $\widehat{\mathbf{M}}_{EL1}$, $\widehat{\mathbf{M}}_{ELP}$, $\widehat{\mathbf{M}}_{EL3}$ and $\widehat{\mathbf{M}}_{EL4}$ are respectively denoted $EL1$, ELP, $EL3$ and $EL4$. As expected, the MSE calculated for each estimators decreases when K increases. The larger the number of data K, the better the precision of estimators. For the EL, the MSE decreases with the difference $n - d$, i.e. the estimation performance increases with the number of prior informations. This motivates the use of EL as soon as prior informations are available. That is why on Figure B.1.a, the MSE of $\widehat{\mathbf{M}}_{EL4}$ (resp. $\widehat{\mathbf{M}}_{EL3}$, $\widehat{\mathbf{M}}_{ELP}$) is lower than the MSE of $\widehat{\mathbf{M}}_{EL3}$ (resp. $\widehat{\mathbf{M}}_{ELP}$, $\widehat{\mathbf{M}}_{EL1}$). Moreover, as seen previously, simulations validate that $\widehat{\mathbf{M}}_{EL1}$ is equal to the SCM.

On the other hand, since the FP estimator is not optimal in Gaussian context, its performance are quite poor. This estimator has been introduced in the literature in cases of non-Gaussian noise models.

Moreover, notice that the MSE of $\widehat{\mathbf{M}}_{EL4}$ (resp. $\widehat{\mathbf{M}}_{EL3}$) reaches the MSE of $\widehat{\mathbf{M}}_{B4}$ (resp. $\widehat{\mathbf{M}}_{B3}$) as soon as the number K of data is large enough: on Figure B.1.a, it approximatively corresponds to $K = 200$. The same conclusion holds for $\widehat{\mathbf{M}}_{ELP}$ and $\widehat{\mathbf{M}}_{B2}$, but for larger K. This can be explained by the fact that both methods exploit the same hypothesis on the structure of the covariance matrix $\mathbf{M}$. For smaller values of K, the information contained in the Gaussian hypothesis plays an effective role which can not be reduced to the observed data. That is why Burg's method is sensitively better than the EL method for small values of K.

Finally, equations (45) and (53) provide closed-form expressions of $\widehat{\mathbf{M}}_{ELP}$ and $\widehat{\mathbf{M}}_{EL3}$ respectively, in opposite to $\widehat{\mathbf{M}}_{B2}$ and $\widehat{\mathbf{M}}_{B3}$ which need an optimization procedure, for the same estimation performance.

This first simulation advocates for the EL method. Despite the fact that it does not use the Gaussian assumption on the data pdf, the EL method shares the same estimation performance as Burg's one, without time consuming for $\widehat{\mathbf{M}}_{ELP}$ and $\widehat{\mathbf{M}}_{EL3}$.

5.1.2 $\rho = 0.5$ *and* $\rho = 0.8$

First comments of Figure B.1.a are still valid for Figure B.1.b and Figure B.1.c. The main difference is that, for $\rho = 0.5$ and even more for $\rho = 0.8$, Burg's method suffers from a relative lack of performance, due to the difficulty of the matrix inversion in its algorithm, for these more correlated situations: indeed, when ρ increases, $\widehat{\mathbf{M}}_{B4}$ deteriorates, in comparison to $\widehat{\mathbf{M}}_{EL4}$. The same comment is valid for $\widehat{\mathbf{M}}_{EL3}$ and $\widehat{\mathbf{M}}_{ELP}$. Figure B.1.c still shows that for $\rho = 0.8$, the MSE of $\widehat{\mathbf{M}}_{B4}$ is even above the MSE of $\widehat{\mathbf{M}}_{EL3}$.

An other important remark is that the distance, in terms of MSE, between $\widehat{\mathbf{M}}_{EL4}$ and $\widehat{\mathbf{M}}_{EL3}$ (resp. $\widehat{\mathbf{M}}_{B4}$ and $\widehat{\mathbf{M}}_{B3}$) seems to be decreasing, i.e. that the supplementary assumption on the particular structure of the Toeplitz matrix brings less additional information for the estimation procedure.

Finally, notice that MSE values seem to be smaller as ρ increases: for instance, for $K = 100$, the MSE of $\widehat{\mathbf{M}}_{EL4}$ is around 0.06 for $\rho = 0.1$, 0.04 for $\rho = 0.5$ and 0.02 for $\rho = 0.8$.

5.1.3 MSE as a function of ρ

To confirm last comment, Figure B.2 presents MSE against ρ for $K = 100$ and $p = 3$.

First comment is that all the MSE decrease as ρ increases. This can be partially explained by the fact that the norm of the matrix $\mathbf{M}$, which is the denominator of the MSE, increases with ρ.

As observed in the previous paragraph, Burg's estimators have a gain in performance smaller than the EL's ones when ρ increases. This is also explained by the fact that Burg's algorithm relies on the inversion of the covariance matrix, which is a difficult issue for large values of ρ. This underlines the robustness of EL. Moreover, $\widehat{\mathbf{M}}_{EL4}$, $\widehat{\mathbf{M}}_{EL3}$ and $\widehat{\mathbf{M}}_{ELP}$ (resp. $\widehat{\mathbf{M}}_{B4}$, $\widehat{\mathbf{M}}_{B3}$ and $\widehat{\mathbf{M}}_{B2}$) approach from each other: this means that the additional information concerning the particular Toeplitz structure of $\mathbf{M}$ (i.e. defined only by ρ), assumed to build $\widehat{\mathbf{M}}_{EL4}$ and $\widehat{\mathbf{M}}_{B4}$, does not allow to improve significantly estimation performance when ρ tends to 1.

To compare the different estimators in a non-Gaussian context, we retain a mixture between the Gaussian distribution and the K-distribution for the data pdf, with a very small shape parameter ν (10^{-3}) to model very impulsive noise. Moreover, this pdf is widely used in signal processing, see for example [13,24,25,36]. The K-distribution is the product of the square root of a random variable g_k which is Gamma distributed, and an independent complex zero-mean Gaussian vector $\mathbf{n}_k$, with covariance matrix $\mathbf{S}$: $\mathbf{x}_k = \sqrt{g}_k \ \mathbf{n}_k$; it is denoted $\mathcal{CK}(\nu, \mathbf{0}, \mathbf{S})$. In order to simulate as better as possible a real situation, the noise is the sum of two independent noises (weighted by a parameter η which deals with the impact of each distribution): a Gaussian noise, which would model thermal noise or interferences, and a K-distribution which would represent an additive non-Gaussian noise (for instance, the clutter in radar context).

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

$$\mathbf{x}_k \sim \frac{1}{\sqrt{2\ (\eta^2 + (1-\eta^2))}} \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}^{1/2} \left[\eta\, \mathcal{CN}\left(\mathbf{0}, \mathbf{I}\right) + (1-\eta)\, \mathcal{CK}\left(\nu, \mathbf{0}, \mathbf{I}\right)\right]. \tag{66}$$

The set of parameters in this subsection is $p = 3$, $K = 100$, $\rho = 0.1$ and $\nu = 10^{-3}$, which means that the K-distribution is extremely impulsive . The value of the correlation coefficient allows to analyze the Burg's method in a suitable context. The simulation results are given in Figure B.3.

For all the estimators except the Fixed Point, the MSE decreases as η increases. In other terms, the performance spoil as the data pdf diverges from the Gaussian one. In the opposite, the Fixed Point performance remains constant as η variates, as it is designed for, see e.g. [36]. Since the FP estimator, like the SCM and EL1, does not assume any particular structure on the covariance matrix $\mathbf{M}$, it makes no sense to compare these estimators to the others. Therefore, ELP and Burg2 have to be considered together while EL3 and Burg3 (resp. EL4 and Burg4) are considered as a third (resp. fourth) set of estimators. As expected, the Fixed Point has the smallest MSE (beyond the three non structured estimators) when the noise is very impulsive, and deteriorates as η increases, to approach the SCM on Gaussian data. Actually, performance of the Fixed Point are almost constant while SCM performs better when η increases.

When η is small enough, data are non-Gaussian and EL method provides significantly better performance than Burg's one. Then, when η becomes close to 1, the noise is approximatively Gaussian and it coincides with results of Figure B.1.a .

Notice that the ratio between the curves $EL3$ and $EL4$ (resp. $Burg3$ and $Burg4$) is constant for each value of η. This ratio corresponds theoretically to a ratio between the expectation of a $\chi^2(3)$ (where 3 is the number of unknown parameters: τ, a_1 and a_2) and the expectation of a $\chi^2(2)$ (where 2 is the number of unknown parameters: τ and ρ). This ratio is equal to 1.5 . This also the case for example, for ELP and $EL3$ (ratio of the expectations of a $\chi^2(4)$ and a $\chi^2(3)$).

5.3 Simulations synthesis

This set of simulations illustrates the use of Empirical Likelihood thanks to an example, the estimation of a structured covariance matrix of an additive noise. It appears that in a Gaussian context and when the data correlation is weak enough, EL competes with the standard method, introduced by Burg in [32] despite the fact that EL does not exploit the a priori pdf of the data. Moreover, simulations show that EL method provides even better performance than the other estimators when the correlation becomes stronger.

Secondly, under non-Gaussian pdf, Burg's estimators performance seem to deteriorate whereas the EL ones remain robust. This is coherent with the fact that EL is a method designed to handle any data distribution.

6 Conclusions

In this paper, an estimation method in unknown noise field, the Empirical Likelihood method, has been introduced. This method does not require an a priori knowledge of the data distribution but it uses different informations like, for instance, moments or specific structure on the parameter of interest. In the field of Signal Processing, such prior information is encountered in many estimation problems, in which the data pdf is not available. Moreover, the theoretical results provided by the EL study shows that under Gaussian assumptions, and without any supplementary information, the Maximum Em-

pirical Likelihood estimator is the same as the Maximum Likelihood estimator.

The second part of this paper has been devoted to a classical and generic example: the estimation of a structured covariance matrix under Gaussian and non-Gaussian assumptions. Under Gaussian assumptions, the EL method has been compared to the classical and structured methods introduced by Burg. It appears that EL performs almost as well as Burg's method. On the other hand, in non-Gaussian context, EL presents good performance even when other considered methods fail.

Moreover, according to prior information used in the EL method, closed-form expressions have been derived. This improves classical methods which generally have to solve an optimization procedure in terms of computational complexity, leading to a substantial gain in time and robustness.

A The Kullback-Leibler divergence

First, in order to class the candidate pdfs in regard of the data, one needs a measure on the pdfs. A natural choice is to use the Kullback-Leibler divergence which allows to compare a candidate Q with the data generating probability P_0. This choice is motivated by the fact that the Maximum Likelihood theory is closely connected to the Kullback-Leibler divergence, see [37]. For Q and P two distinct pdfs, the Kullback-Leibler divergence is defined as follows:

$$\mathcal{K}(Q,P) = \begin{cases} -\int \log\left(\frac{dQ}{dP}\right) dP & \text{if } Q \text{ is absolutely continuous with respect to } P \\ +\infty & \text{otherwise.} \end{cases} \tag{A.1}$$

This gives a feasible measure $\mathcal{K}(Q, P_K)$. At this point, it is interesting to notice that for any Q which is not absolutely continuous with respect to P_K, the divergence diverges. Therefore, the only convenient pdfs are those which are absolutely continuous with respect to P_K.

B Proof of Theorem 3.1

The main technical difficulty, evaluating the empirical likelihood $EL(\boldsymbol{\theta})$ at any given $\boldsymbol{\theta}$, concerns the constrained problem optimization which is solved here by a Lagrangian method.

Lemma B.1 (Explicit expression for $EL(\boldsymbol{\theta})$)
If the assumptions of Theorem 3.1 hold, then the supremum appearing in the definition of $EL(\boldsymbol{\theta})$ is realized and, there exists λ^* such as the optimal weights are given by

$$q_k^* = \frac{1}{K}\left(1 + \lambda^{*\top}\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta})\right)^{-1}.$$

$EL(\boldsymbol{\theta})$ is then given by

$$EL(\boldsymbol{\theta}) = \min_{\lambda}\left\{\prod_{k=1}^{K}\left(K(1 + \lambda^{\top}\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}))\right)^{-1}\right\}$$

***Proof* B.0.1**
First, note that $-\log$ being a decreasing function,

$$-\log\left(EL(\boldsymbol{\theta})\right) = -\log \sup_{(q_1,\ldots,q_K)}\left\{\prod_{k=1}^{K} q_k \middle| \sum_{k=1}^{K} q_k\,\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}) = \mathbf{0},\ \sum_{k=1}^{K}(q_k - 1/K) = 0\right\} \tag{B.1}$$

$$= \min_{(q_1,\ldots,q_K)}\left\{-\log\left(\prod_{k=1}^{K} q_k\right) \middle| \sum_{k=1}^{K} q_k\,\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}) = \mathbf{0},\ \sum_{k=1}^{K}(q_k - 1/K) = 0\right\} \tag{B.2}$$

$$= 2\min_{(q_1,\ldots,q_K,\lambda,\gamma)}\left\{-\sum_{k=1}^{K}\log(q_k) + K\,\lambda^{\top}\sum_{k=1}^{K} q_k\,\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}) - \gamma\sum_{k=1}^{K}(q_k - 1/K)\right\}. \tag{B.3}$$

The first order condition gives

$$-1/q_k^* + K\,\lambda^{*\top}\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}) - \gamma^* = 0.$$

Multiplying by q_k^* and summing over k, one gets $\gamma^* = -K$ and then

$$q_k^* = \frac{1}{K}\left(1 + \lambda^{*\top}\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta})\right)^{-1}.$$

Finally,

$$EL(\boldsymbol{\theta}) = \min_{\lambda}\left\{2\sum_{k=1}^{K}\log\left(K(1 + \lambda^{\top}\mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}))\right)\right\}.$$

Now we can start the proof of Theorem 3.1. In the specific case where n, the dimension of arrival space of function $\mathbf{m}$ is equal to the dimension d of the parameter of interest and under mild conditions, the optimum is reached at the $\boldsymbol{\theta}$ such as $\lambda^* = \mathbf{0}$. Therefore, the optimal weights are

$$\widehat{q_k^*} = \frac{1}{K}\left(1 + \mathbf{0}^\top \mathbf{m}(\mathbf{x}_k, \hat{\boldsymbol{\theta}}_{MEL})\right)^{-1} = \frac{1}{K}\,, \tag{B.4}$$

and thus

$$\sup_{\boldsymbol{\theta}} \{EL(\boldsymbol{\theta})\} = K^{-K}\,. \tag{B.5}$$

Therefore, the log-likelihood ratio writes:

$$\begin{aligned}
ELR(\boldsymbol{\theta}) &= -2\log\left(\frac{EL(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}}\{EL(\boldsymbol{\theta})\}}\right) && \text{(B.6)}\\
&= -2\log\left(K^K \sup_{(q_1,\ldots,q_K)} \left\{ \prod_{k=1}^{K} g_{(\boldsymbol{\theta},q_1,\ldots,q_K)}(\mathbf{x}_k) \middle| E_G[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})] = \mathbf{0} \right\}\right) && \text{(B.7)}\\
&= 2 \min_{(q_1,\ldots,q_K)} \left\{ -\sum_{k=1}^{K} \log(Kq_k) \middle| E_G[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})] = \mathbf{0} \right\} && \text{(B.8)}\\
&= 2\,K \min_{(q_1,\ldots,q_K)} \left\{ -\frac{1}{K}\sum_{k=1}^{K} \log\left(\frac{q_k}{1/K}\right) \middle| E_G[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})] = \mathbf{0} \right\} && \text{(B.9)}\\
&= 2\,K \min_{(q_1,\ldots,q_K)} \left\{ -\int \log\left(\frac{dG}{dP_K}\right) dP_K \middle| E_G[\mathbf{m}(\mathbf{x},\boldsymbol{\theta})] = \mathbf{0} \right\} && \text{(B.10)}\\
&= \mathcal{C}(\boldsymbol{\theta}), && \text{(B.11)}
\end{aligned}$$

which concludes the proof.

Now, to derive $\hat{\boldsymbol{\theta}}_{MEL}$, one writes

$$\arg\min_{\boldsymbol{\theta}} \{ELR(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \min_{\lambda} \left\{ 2\sum_{k=1}^{K} \log\left(1 + \lambda^\top \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta})\right) \right\}. \tag{B.12}$$

References

[1] S. M. Kay, Fundamentals of statistical signal processing - Detection theory, Vol. 2, Prentice-Hall PTR, 1998.

[2] H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, II and III, John Wiley & Sons, New York, 1968, 1971.

[3] L. L. Scharf, D. W. Lytle, Signal detection in gaussian noise of unknown level: an invariance application, IEEE Trans.-IT 17 (1971) 404–411.

[4] S. Haykin, Array Signal Processing, Prentice-Hall Signal Processing Series, Englewood Cliffs, New Jersey, 1985.

[5] H. L. Van Trees, Detection, Estimation and Modulation Theory, Part IV: Optimum Array Processing, John Wiley & Sons, New York, 2002.

[6] E. J. Kelly, An adaptive detection algorithm, IEEE Trans.-AES 23 (1) (1986) 115–127.

[7] S. Kraut, L. L. Scharf, L. T. Mc Whorter, Adaptive subspace detectors, IEEE Trans.-SP 49 (1) (2001) 1–16.

[8] J. G. Proakis, Digital Communications, McGraw-Hill, Third Ed., New York, 1995.

[9] M. Rangaswamy, J. H. Michels, D. D. Weiner, Multichannel detection for correlated non-gaussian random processes based on innovations, IEEE Trans.-SP 43 (8) (1995) 1915–1922.

[10] J.-F. Cardoso, Source separation using higher order moments, in: Proc. IEEE-ICASSP, Glasgow, 1989, pp. 2109–2112.

[11] P. M. Djuric et al., Particle filtering, IEEE SP Magazine 20 (5) (2003) 19–38.

[12] J. B. Billingsley, Ground clutter measurements for surface-sited radar, Tech. Rep. 780, MIT (February 1993).

[13] S. Watts, Radar detection prediction in sea clutter using the compound k-distribution model, IEE Proceeding, Part. F 132 (7) (1985) 613–620.

[14] T. Nohara, S. Haykin, Canada east coast trials and the k-distribution, IEE Proceeding, Part. F 138 (2) (1991) 82–88.

[15] A. Farina, A. Russo, F. Scannapieco, Radar detection in coherent weibull clutter, IEEE Trans.-ASSP 35 (6) (1987) 893–895.

[16] A. B. Owen, Empirical likelihood ratio confidence regions, Annals of Statistics 18 (1990) 90–120.

[17] A. Dogandzic, B. Zhang, Distributed estimation and detection for sensor networks using hidden markov random field models, IEEE Trans.-SP 54 (8) (2006) 3200–3215.

[18] F. Pascal, J.-P. Barbot, H. Harari-Kermadec, R. Suyama, P. Larzabal, An empirical likelihood method for data aided channel identification in unknown noise field, in: Proc. of the European Signal Processing Conf., EUSIPCO-08, Lausanne, Switzerland, 2008.

[19] H. Harari-Kermadec, F. Pascal, On the use of empirical likelihood for non-gaussian clutter covariance matrix estimation, in: Proc. of the IEEE-RADAR-08, Roma, Italy, 2008.

[20] P. J. Huber, The 1972 wald lecture robust statistics : A review, Annals of Mathematical Statistics 43 (4) (1972) 1041–1067.

[21] G. Pailloux, P. Forster, J.-P. Ovarlez, F. Pascal, On persymmetric covariance matrices in adaptive detection, in: Proc. IEEE-ICASSP, Philadelphia, 2008.

[22] U. Grenander, G. Szego, Toeplitz Forms and Their Applications, University of California Press, Berkeley and Los Angeles, 1958.

[23] D. S. G. Pollock, A Handbook of Time-Series Analysis, Signal Processing and Dynamics, Academic Press, 1999.

[24] E. Conte, M. Lops, G. Ricci, Adaptive detection schemes in compound-gaussian clutter, IEEE Trans.-AES 34 (4) (1998) 1058–1069.

[25] F. Gini, M. V. Greco, Sub-optimum approach to adaptive coherent radar detection in compound-gaussian clutter, IEEE Trans.-AES 35 (3) (1999) 1095–1103.

[26] R. M. Gray, Toeplitz and circulant matrices, Tech. Rep. 6504-1, Stanford University Information Systems Laboratory (April 1977).
URL http://www-isl.stanford.edu/$\sim$gray/compression.html

[27] F. Pascal, P. Forster, J.-P. Ovarlez, P. Larzabal, Performance analysis of covariance matrix estimates in impulsive noise, IEEE Trans.-SP.

[28] A. B. Owen, Empirical Likelihood, Chapman & Hall/CRC, Boca Raton, 2001.

[29] Y. S. Qin, J. Lawless, Empirical likelihood and general estimating equations, Annals of Statistics 22 (1) (1994) 300–325.

[30] L. Cai, H. Wang, A Persymmetric Multiband GLR Algorithm, IEEE trans. on Aerosp. Electron. System (1992) 806–816.

[31] E. Conte, A. D. Maio, Exploiting Persymmetry for CFAR Detection in Compound-Gaussian Clutter, IEEE trans. on Aerosp. Electron. System 39 (2003) 719–724.

[32] J. P. Burg, D. G. Luenberger, D. L. Wenger, Estimation of structured covariance matrices, Proc. IEEE 70 (9) (1982) 963–974.

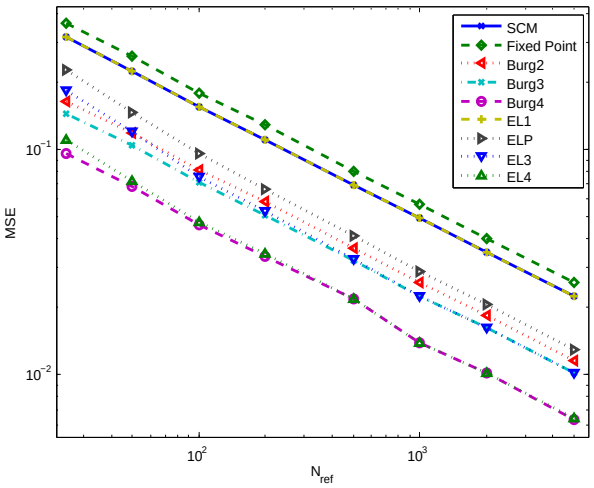
[33] D. R. Fuhrmann, Application of toeplitz covariance estimation to adaptive beamforming and detection, IEEE Trans.-SP 39 (10) (1991) 2194–2198.

[34] E. Conte, A. De Maio, G. Ricci, Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive cfar detection, IEEE Trans.-SP 50 (8) (2002) 1908–1915.

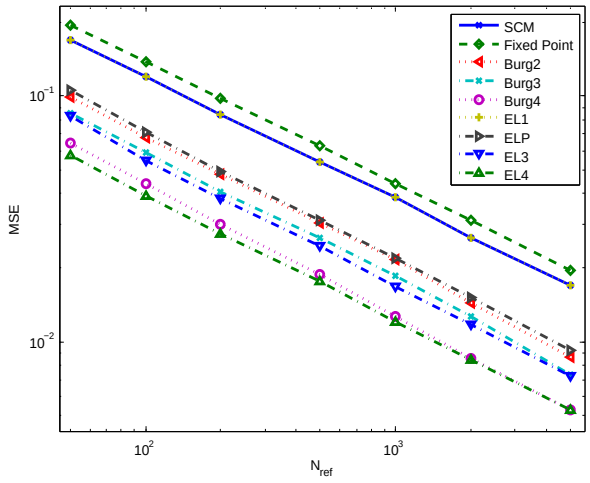
[35] F. Gini, M. V. Greco, Covariance matrix estimation for cfar detection in correlated heavy tailed clutter, Signal Processing, special section on SP with Heavy Tailed Distributions 82 (12) (2002) 1847–1859.

[36] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, P. Larzabal, Covariance structure maximum likelihood estimates in compound gaussian noise : Existence and algorithm analysis, IEEE Trans.-SP 56 (1) (2008) 34–48.

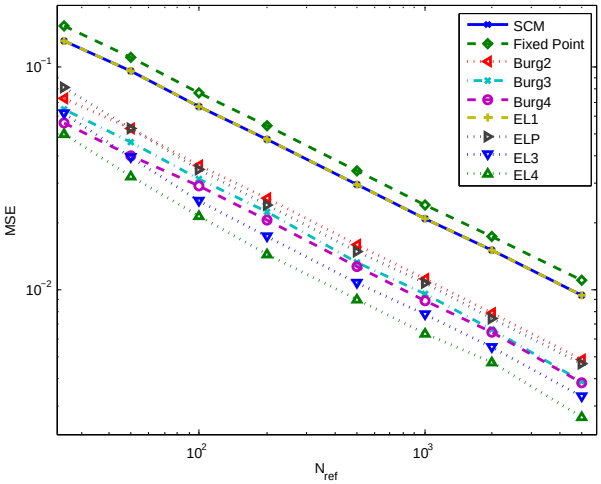
[37] A. W. van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.



(a) $\rho = 0.1$



(b) $\rho = 0.5$



(c) $\rho = 0.8$

Fig. B.1. Complex Gaussian case: MSE of each estimator against the number N_ref of data, for different correlation coefficient ρ and for $p = 3$ and $\tau = 1$.

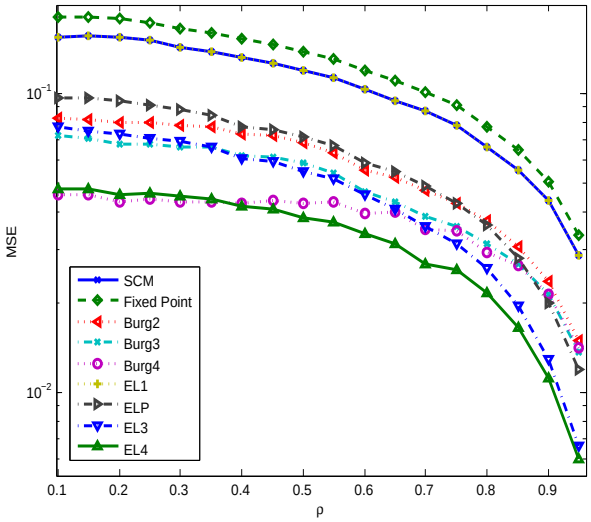


Fig. B.2. MSE against the correlation coefficient ρ, for $K = 100$, $p = 3$ and complex Gaussian data and for the different estimation procedures.

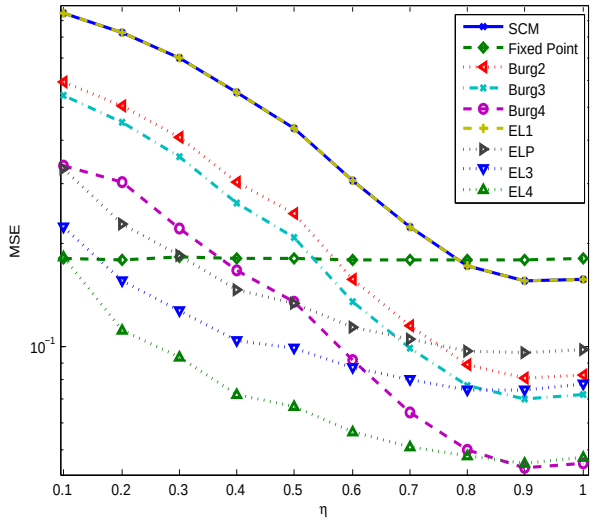


Fig. B.3. MSE against η, the shape parameter of the K-Distribution, for the different estimation procedures and for $\rho = 0.1$, $p = 3$ and $K = 100$.